4/5/1 (Item 1 from file: 351)
DIALOG(R)File 351:Derwent WPI

012386060 **Image available**
WPI Acc No: 1999-192167/199917
XRPX Acc No: N99-140742
Large capacity multi-class core ATM switch architecture
Patent Assignee: NEC CORP (NIDE); NEC USA INC (NIDE)
Inventor: FAN R; MARK B; RAMAMURTHY G; MARK B L
Number of Countries: 027 Number of Patents: 003
Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week	
EP 901302	A2	19990310	EP 98107930	A	19980430	199917	B
JP 11088374	A	19990330	JP 9892088	A	19980403	199923	
US 6324165	B1	20011127	US 97923978	A	19970905	200175	

Priority Applications (No Type Date): US 97923978 A 19970905
Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
EP 901302	A2	E	43	H04Q-011/04	
Designated States (Regional): AL AT BE CH CY DE DK ES FI FR GB GR IE IT LI LT LU LV MC MK NL PT RO SE SI					
JP 11088374	A		29	H04L-012/28	
US 6324165	B1			H04L-012/26	

Abstract (Basic): EP 901302 A2
NOVELTY - The large capacity triple-buffered switch is preferably a single stage switch which may be used as a switching element in a still larger capacity multi-stage switch. The core switch module (10) consists of 16 input ports (IP), 16 output ports (OP) and one multicast output port (MOP), all connected to a high speed terminal bus (20). The core module interconnects the input/output modules to large buffers and intelligent/management mechanisms
DETAILED DESCRIPTION - INDEPENDENT CLAIMS are included for an ATM switch scheduler and a buffer
USE - For servicing requests originated from various classes of sources in ATM network.
ADVANTAGE - Can flexibly vary bandwidth of switch and cope with multiple traffic class requirements
DESCRIPTION OF DRAWING(S) - The drawing is a diagram depicting architecture of core switch module according to preferred embodiment of present invention.
Core switch module (10)
Input ports (IP)
Output ports (OP)
Multicast output port (MOP)
High speed terminal bus (20)
pp; 43 DwgNo 1/18
Title Terms: CAPACITY; MULTI; CLASS; CORE; ATM; SWITCH; ARCHITECTURE
Derwent Class: W01
International Patent Class (Main): H04L-012/26; H04L-012/28; H04Q-011/04
International Patent Class (Additional): H04L-012/56; H04Q-003/00
File Segment: EPI

4/5/2 (Item 1 from file: 347)
DIALOG(R)File 347:JAPIO

06146834 **Image available**
LARGE CAPACITY MULTI-CLASS CORE ATM SWITCH ARCHITECTURE

PUB. NO.: 11-088374 A]
PUBLISHED: March 30, 1999 (19990330)
INVENTOR(s): FAN RUIXUE

MARK BRIAN
RAMAMURTHY GOPALAKRISHNAN
APPLICANT(s): NEC CORP
APPL. NO.: 10-092088 [JP 9892088]
FILED: April 03, 1998 (19980403)
PRIORITY: 923978 [US 923978], US (United States of America), September 05, 1997 (19970905)
INTL CLASS: H04L-012/28; H04Q-003/00

ABSTRACT

PROBLEM TO BE SOLVED: To provide a large capacity ATM core switch structure for supporting plural traffic classes and service quality guarantee.

SOLUTION: A switch efficiently adjusts real-time and non-real-time multi-cast flows and the switch is constituted of a high-speed core module 10 for mutually connecting input/output modules provided with a large capacity buffer and an intelligent scheduling/buffer management mechanism. Scheduling is realized by using new dynamic rate control for controlling internal congestion and achieving fair throughput performance between the flows competing at a switch bottleneck part. In a dynamic rate control system, the flow is rate-controlled corresponding to congestion information monitored at the bottleneck part inside the switch. For respective switch flows, a rate for which a dynamic rate component for fairly dividing an unused band width is added to a minimum service rate is guaranteed.

(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号

特開平11－88374

(43)公開日　平成11年(1999)３月30日

(51)Int.Cl.[6]	識別記号	ＦＩ	
Ｈ０４Ｌ　12/28		Ｈ０４Ｌ　11/20	Ｇ
Ｈ０４Ｑ　3/00		Ｈ０４Ｑ　3/00	
		Ｈ０４Ｌ　11/20	Ｈ

審査請求　有　　請求項の数27　ＯＬ　（全 29 頁）

(21)出願番号　特願平10－92088

(22)出願日　平成10年(1998)４月３日

(31)優先権主張番号　０８／９２３９７８

(32)優先日　1997年９月５日

(33)優先権主張国　米国（ＵＳ）

(71)出願人　000004237
日本電気株式会社
東京都港区芝五丁目７番１号

(72)発明者　ルイクシュー　ファン
アメリカ合衆国，ニュージャージー　08540，プリンストン，４　インディペンデンス　ウエイ，エヌ・イー・シー・ユー・エス・エー・インク内

(74)代理人　弁理士　後藤　洋介　（外１名）

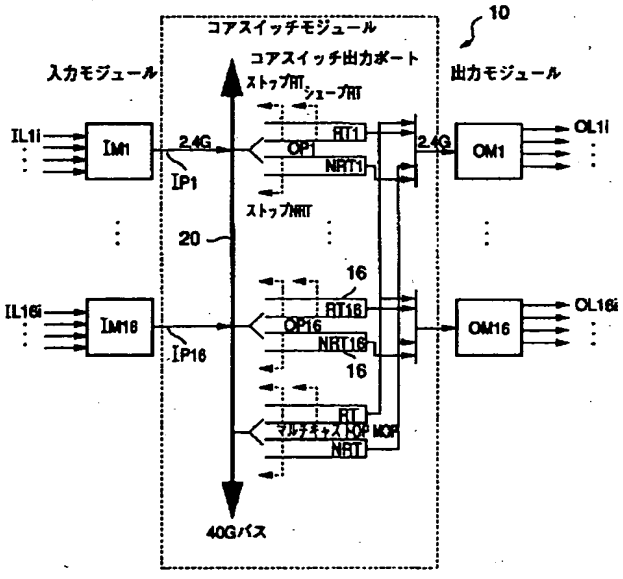
最終頁に続く

(54)【発明の名称】　大容量マルチクラスコアＡＴＭスイッチアーキテクチャ

(57)【要約】　（修正有）

【課題】　複数のトラフィッククラス及びサービス品質保証をサポートする大容量ＡＴＭコアスイッチ構造を提供する。

【解決手段】　スイッチは、リアルタイム及びノンリアルタイムマルチキャストフローを効率的に調整する。スイッチは、大容量バッファを備えた入力／出力モジュールを相互接続する高速コアモジュールと、インテリジェントスケジューリング／バッファ管理メカニズムとで構成される。スケジューリングは、内部輻輳を制御し、スイッチボトルネック部で競合するフロー間に公平なスループット性能を達成する新規な動的レート制御を用いて実現できる。動的レート制御方式では、フローは、スイッチ内のボトルネック部で監視される輻輳情報に応じてレート制御される。各スイッチフローには、最低サービスレートに、未使用帯域幅を公平に配分する動的レート成分を加えたレートが、保証される。



1

【特許請求の範囲】

【請求項１】　コアスイッチと、上記コアスイッチの入力側に接続された複数個の入力モジュールと、上記コアスイッチの出力側に接続された複数個の出力モジュールとを備え、

上記コアスイッチは、

ＴＤＭバスと、

上記ＴＤＭバスに接続された複数個の入力ポートと、

上記ＴＤＭバスに接続された複数個の出力バッファと、

上記出力バッファにそれぞれ接続された複数個の出力ポートと、

上記出力ポートの各々に接続されたマルチキャスト出力バッファとを備え、

上記各入力モジュールは、

上記出力ポートの数に対応して設けられ、且つ、それぞれ複数個の入力バッファを有する出力ポートプレーンと、

上記入力バッファにおいてセルをスケジューリングするための入力モジュールスケジューラとを備え、

上記出力モジュールの各々は、

複数個の出力回線プレーンであって、各々が、出力回線に連結された複数個の出力回線バッファを有する出力回線プレーンと、

上記出力バッファにおいてセルをスケジューリングするための出力モジュールスケジューラとを備えた、異なるサービス品質要求を有する異なるクラスのストリームをサポートできるＡＴＭスイッチ。

【請求項２】　請求項１において、上記コアスイッチの上記出力バッファの各々は、各出力バッファ内のセルのレベルが第１のしきい値に到達するとシェーピング信号を発生し、各出力バッファ内のセルのレベルが第２のしきい値に到達するとストップ信号を発生するオーバーフロー制御部を備えたことを特徴とするＡＴＭスイッチ。

【請求項３】　請求項１において、上記複数個の入力バッファ内の各セルに対してキュータイムスタンプを付与するコネクション受付制御部と、

現在時刻を与えるタイマとをさらに備え、

上記入力モジュールスケジューラは、さらに、上記キュータイムスタンプと上記現在時刻を比較し、現在時刻に等しいキュータイムスタンプを有する上記セルの各々をスケジューリングするコンパレータを含むことを特徴とするＡＴＭスイッチ。

【請求項４】　請求項１において、上記コアスイッチの上記複数個の出力バッファとして、リアルタイム出力バッファが設けられると共に、上記コアスイッチは、更に、

上記リアルタイム出力バッファの数に対応した数だけ設けられ、上記出力ポートの一つにそれぞれ接続された複数個のノンリアルタイム出力バッファと、

上記出力ポートの各々に接続されたノンリアルタイムマルチキャストバッファとを備えていることを特徴とするＡＴＭスイッチ。

2

【請求項５】　請求項３において、上記入力モジュールスケジューラ及び上記出力モジュールスケジューラの各々は、

上記キュータイムスタンプを記憶し、上記コンパレータに上記キュータイムスタンプを提供するタイムスタンプ記憶装置と、

サービス資格のあるセルの伝送レートをシェーピングするための、上記コンパレータに接続された仮想レートシェーピング装置と、

サービス資格のあるセルの数をカウントするための、複数個の仮想キューカウンタと、

サービス資格のあるセルをスケジューリングするためのサービススケジューリング装置と、

上記タイムスタンプを動的に更新するための計算エンジンとをさらに備えたことを特徴とするＡＴＭスイッチ。

【請求項６】　請求項５において、上記出力モジュールから上記入力モジュールへのレートフィードバックをさらに備えたことを特徴とするＡＴＭスイッチ。

【請求項７】　コアスイッチと、上記コアスイッチの入力側に接続された複数個の入力モジュールと、上記コアスイッチの出力側に接続された複数個の出力モジュールとを備え、

上記コアスイッチは、

ＴＤＭバスと、

上記ＴＤＭバスに接続された複数個の入力ポートと、

上記ＴＤＭバスに接続された複数個の出力バッファと、

上記出力バッファにそれぞれ接続された複数個の出力ポートと、

上記出力ポートの各々に接続されたマルチキャスト出力バッファとを備え、

上記出力モジュールの各々は、複数個の出力回線を有し、且つ、

複数個の出力回線プレーンであって、各々が、上記出力回線に連結された複数個の出力回線バッファを有する出力回線プレーンと、

上記出力バッファにおいてセルをスケジューリングするための出力モジュールスケジューラとを備え、

上記各入力モジュールは、

上記出力ポートの数に対応して設けられた複数個の出力ポートプレーンであって、それぞれ、上記出力回線の数に対応する複数個の出力回線プレーンを有し、上記出力回線プレーンの各々が複数個の入力バッファを有する出力ポートプレーンと、

上記入力バッファにおいてセルをスケジューリングするための入力モジュールスケジューラとを備えた、異なるサービス品質要求を有する異なるクラスのストリームをサポートできるＡＴＭスイッチ。

【請求項８】　請求項７において、上記コアスイッチの

上記出力バッファの各々は、各出力バッファ内のセルのレベルが第１のしきい値に到達するとシェーピング信号を発生し、各出力バッファ内のセルのレベルが第２のしきい値に到達するとストップ信号を発生するオーバーフロー制御部を備えたことを特徴とするＡＴＭスイッチ。

【請求項９】　請求項７において、上記複数個の入力バッファ内の各セルに対してキュータイムスタンプを付与するコネクション受付制御部と、

現在時刻を与えるタイマとをさらに備え、

上記入力モジュールスケジューラは、さらに、上記キュータイムスタンプと上記現在時刻を比較し、現在時刻に等しいキュータイムスタンプを有する上記セルの各々をスケジューリングするコンパレータを含むことを特徴とするＡＴＭスイッチ。

【請求項１０】　請求項７において、上記コアスイッチの上記複数個の出力バッファとして、リアルタイム出力バッファが設けられると共に、上記コアスイッチは、更に、

上記リアルタイム出力バッファの数に対応した数だけ設けられ、上記出力ポートの一つにそれぞれ接続された複数個のノンリアルタイム出力バッファと、

上記出力ポートの各々に接続されたノンリアルタイムマルチキャストバッファとを備えたことを特徴とするＡＴＭスイッチ。

【請求項１１】　請求項９において、上記入力モジュールスケジューラ及び上記出力モジュールスケジューラの各々は、

上記キュータイムスタンプを記憶し、上記コンパレータに上記キュータイムスタンプを提供するタイムスタンプ記憶装置と、

サービス資格のあるセルの伝送レートをシェーピングするための、上記コンパレータに接続された仮想レートシェーピング装置と、

サービス資格のあるセルの数をカウントするための、複数個の仮想キューカウンタと、

サービス資格のあるセルをスケジューリングするためのサービススケジューリング装置と、

上記タイムスタンプを動的に更新するための計算エンジンとをさらに備えたことを特徴とするＡＴＭスイッチ。

【請求項１２】　請求項１１において、記載の上記出力モジュールから上記入力モジュールへのレートフィードバックをさらに備えたことを特徴とするＡＴＭスイッチ。

【請求項１３】　複数個の入力ポート及び複数個の出力ポートに接続された中央バスを有するコアスイッチと、

上記出力ポートに接続された複数個の出力モジュールと、

上記入力ポートに接続された複数個の入力モジュールとを備え、当該各入力モジュールは、複数個の共通のキューと、スケジューラ部と、コネクション受付制御部とを

備えた入力モジュールとを含み、

上記コネクション受付制御部は、異なるクラスに上記キューを割り当て、入力セルストリームを、上記セルストリームに対して要求されたサービス品質にしたがって、適切なキューにルーティングすることを特徴とする、異なるサービス品質要求を有する異なるクラスのセルストリームをサポートできるＡＴＭスイッチ。

【請求項１４】　コアスイッチと、上記コアスイッチの入力側に接続された複数個の入力モジュールとを備え、

上記コアスイッチは、

ＴＤＭバスと、

上記ＴＤＭバスに接続された複数個の入力ポートと、

上記ＴＤＭバスに接続された複数個の出力バッファと、

上記出力バッファにそれぞれ接続された複数個の出力ポートとを備え、

上記入力モジュールの各々は、

複数個の入力バッファと、

上記入力バッファの各々に対して最低保証レート及び超過配分加重を割り当てるコネクション受付制御部と、

上記最低保証レート及び利用可能な未使用帯域幅の割当から構成されるレートにしたがって、上記入力バッファ内のセルをスケジューリングするための入力モジュールスケジューラとを含み、上記配分は過剰配分重みに比例することを特徴とする、異なるサービス品質要求を有する異なるクラスのストリームをサポートできるＡＴＭスイッチ。

【請求項１５】　請求項１４において、上記出力バッファのセル占有率にしたがって、利用可能な未使用帯域幅を調整するための閉ループ制御器をさらに備えたことを特徴とするＡＴＭスイッチ。

【請求項１６】　請求項１５において、記載の上記出力バッファのうちのどれか１つのバッファのセル占有率が所定のシェーピングしきい値に到達するたびに、上記出力バッファのいずれかからシェピングレート信号を送出するオーバーロード制御部をさらに備え、上記スケジューラは上記シェーピング信号を受け、保証された最小レートのみにしたがって上記入力バッファをスケジューリングすることを特徴とするＡＴＭスイッチ。

【請求項１７】　入力セルストリームのキューを作るための複数個の入力バッファと、

上記入力バッファから上記セルストリームのキューセルを受取り、上記キューセルを各出力ポートに与える複数個のコアバッファと、

上記出力ポートから上記キューセルを受取り、上記キューセルを各出力回線に与える複数個の出力バッファと、

上記入力バッファからの上記キューセルの送出をスケジューリングするためのスケジューラと、

上記スケジューラに上記コアバッファの負荷情報を伝送するための、上記コアバッファから上記スケジューラへの第１のフィードバックループと、

上記スケジューラに上記出力バッファの負荷情報を伝送するための、上記出力バッファから上記スケジューラへの第２のフィードバックループとを有し、
上記スケジューラは、上記第１及び上記第２のフィードバックループから受けとった情報にしたがって、上記入力モジュールからの上記キューセルの送出をスケジューリングすることを特徴とする、異なるサービス品質要求を有する異なるクラスのストリームをサポートできる、３重バッファ付ＡＴＭスイッチ。

【請求項１８】　請求項１７において、上記セルストリームに最低保証レートを割り当てるためのコネクション受付制御部をさらに備え、上記スケジューラは、上記最低保証レート以上のレートで、上記入力バッファにおいて上記キューセルをスケジューリングすることを特徴とする３重バッファ付ＡＴＭスイッチ。

【請求項１９】　請求項１８において、上記スケジューラは、動的レートにしたがって上記入力バッファにおいて上記キューセルをスケジューリングし、上記動的レートは、上記最低保証レートに、上記第１及び上記第２のフィードバックループの情報から決定された未使用帯域幅の割当を加えてなることを特徴とする３重バッファ付ＡＴＭスイッチ。

【請求項２０】　請求項１９において、上記コネクション受付制御部はさらに、上記キューの各々にキュータイムスタンプを割り当て、上記スケジューラは、上記キュータイムスタンプと現在時刻を比較し、現在時刻以下のキュータイムスタンプを有する各キューをサービス資格ありと定義するコンパレータをさらに備えたことを特徴とする３重バッファ付ＡＴＭスイッチ。

【請求項２１】　スケジューリングすべきキュー内のセルに割り当てられたセルタイムスタンプを記憶するための第１のメモリと、
実際のキュー負荷を記憶するための第２のメモリと、
仮想キュー内のセルを記憶するための第３のメモリと、
現在時刻を発生する現在時刻発生器と、
上記セルタイムスタンプと現在時刻を比較し、現在時刻以下のセルタイムスタンプを有するセルをサービス資格ありと定めるための複数個のコンパレータと、
上記サービス資格のあるセルから上記仮想キューにセルを割り当てるための仮想レートセレクタと、
上記仮想キューからサービス対象のキューを選択するサービススケジューリングセレクタと、
上記第１メモリ内のセルをスケジューリング及び再スケジューリングするための計算エンジンとを備えたＡＴＭスイッチのためのスケジューラ。

【請求項２２】　バッファであって、上記バッファにおける負荷レベルを監視し、上記負荷レベルが第１のしきい値に到達すると、上記バッファへの入力を最小レベルに低減せるようにシェーピング信号を発生し、さらに、上記負荷レベルが第２のしきい値に到達すると、上記バッファへのあらゆる入力を停止するようにストップ信号を発生するための第１の監視回路を有するバッファ。

【請求項２３】　請求項１４において、上記バッファ上で利用可能な未使用帯域幅を推定し、上記推定を示す信号を発生するための推定回路をさらに備えたバッファ。

【請求項２４】　スケジューリングすべきキューに割り当てられたキュータイムスタンプを記憶するための第１のメモリと、
実際のキュー負荷を記憶するための第２のメモリと、
仮想キュー内のセルを記憶するための第３のメモリと、
現在時刻を発生する現在時刻発生器と、
上記キュータイムスタンプと現在時刻を比較し、現在タイムスタンプ以下のセルタイムスタンプを有するキューをサービス資格ありと定めるための複数個のコンパレータと、
上記サービス資格のあるセルから上記仮想キューにセルを割り当てるための仮想レートセレクタと、
上記仮想キューからサービス対象のキューを選択するサービススケジューリングセレクタと、
上記第１メモリ内のセルをスケジューリング及び再スケジューリングするための計算エンジンとを備えたＡＴＭスイッチのためのスケジューラ。

【請求項２５】互いにレートの異なるデータを伝送する複数のクラスのセルストリームをスイッチするＡＴＭスイッチにおいて、入力側及び出力側を備えたコアスイッチと、前記コアスイッチの入力側に接続された少なくとも一つの入力モジュールと、前記コアスイッチの出力側及び出力回線に接続された少なくとも一つの出力モジュールとを備え、前記出力モジュールから前記入力モジュールに対して、前記出力回線上の送信レートをフィードバックする少なくとも一つのフィードバックループが設けられていることを特徴とするＡＴＭスイッチ。

【請求項２６】請求項２５において、前記入力及び前記出力モジュールには、それぞれスケジューラが設けられており、前記各スケジューラでは、前記各クラスのセルストリームの最低保証レートが、維持されるように、スケジューリングすることを特徴とするＡＴＭスイッチ。

【請求項２７】請求項２６において、前記各スケジューラは、前記最低保証レートに加えて、各クラスの優先度に応じて定められた割合で、超過レートを割り当てることを特徴とするＡＴＭスイッチ。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、非同期転送モード（ＡＴＭ）ネットワークに関し、特に、ＡＴＭフォーラムにより定義されている種々のクラスにおけるソースからの要求に効率的に応え得る大容量マルチクラスコアＡＴＭスイッチに関する。

【０００２】

【従来の技術】過去、電話ネットワークとコンピュータ

ネットワークは、異なる分野で開発されてきた。効果的なリアルタイム通信を保証するために、電話ＴＤＭ(Time Division Multiplexing)ネットワークでは、呼が継続している間維持されるチャネルが設定される。一方、コンピュータネットワーク上で転送されるデータの多くはリアルタイムデータではないため、チャネルを設定することなく、パケットをルーティングするパケット交換が行われている。

【０００３】ＴＤＭネットワークにおける一つの問題点は、不使用状態にあるソースが設定されたチャネルを不必要に占有してしまうことである。他方、パケット交換に伴なう問題点は、パケット交換では高いプロトコルオーバーヘッドを必要とするため、リアルタイム通信に適さないということである。

【０００４】非同期転送モード（ＡＴＭ）技術は、将来の通信交換および伝送インフラストラクチャのための重要な技術として登場してきた。尚、ＡＴＭネットワークに関する記事を集めたものとして、「Ｌｅｃｔｕｒｅ Ｎｏｔｅｓ ｉｎ Ｃｏｍｐｕｔｅｒ Ｓｃｉｅｎｃｅ」（Ｂｒｏａｄｂａｎｄ Ｎｅｔｗｏｒｋ Ｔｅｌｅｔｒａｆｆｉｃ、Ｊａｍｅｓ Ｒｏｂｅｒｔｓ、Ｕｇｏ Ｍｏｃｃｉ、Ｊｏｒｍａ Ｖｉｒｔａｍｏ（編者）、Ｖｏｌ．１１５５、Ｓｐｒｉｎｇｅｒ １９９１年、ＩＳＢＮ ３－５４０－６１８１５－５）がある。

【０００５】ＡＴＭネットワークの主な長所は、互いに異なる種々のトラフィック特性及びサービス品質（ＱｏＳ）要求をともなうアプリケーションをサポートできる点にある。ＡＴＭネットワークの目標は、ＴＤＭネットワークとパケット交換の長所を組み合わせて活用するとともに、これらネットワークのそれぞれの短所を取り除くことである。すなわち、ＡＴＭ交換は、ＴＤＭネットワーク及びパケット交換ネットワークに取って代わるネットワークを提供することができる。

【０００６】ＡＴＭフォーラムはＡＴＭ設計のために種々のガイドラインを制定してきた。これらのガイドラインは、ＡＴＭフォーラムの各種刊行物において見ることができる。ここでは、説明の便宜上、関連するいくつかの指針及び略号を以下に説明する。

【０００７】現在、ＡＴＭフォーラムは、４つの主要なトラフィックサービスクラス、即ち、固定ビットレート（ＣＢＲ）、可変ビットレート（ＶＢＲ）、利用可能なビットレート（ＡＢＲ）、及び無指定ビットレート（ＵＢＲ）のサービスクラスを定めており、これらは、リアルタイムトラフィック及びノンリアルタイムトラフィックのためのサービスクラスに分類される。ＣＢＲサービスはリアルタイムトラフィックに使用され、特に、音声トラフィックに用いられる。ＣＢＲ呼要求の受付の可否は、要求されたピークレートにより決定することができる。ＶＢＲサービスは映像伝送に使用される。このサービスは非常にバースト性が高いトラフィックを対象としており、ＶＢＲ呼要求の受付の際には、ピークレートと、持続レートと、バーストサイズとが考慮される。受付けた際にも、リーキバケットを用いるなどして、このようなソースからの伝送を調整することが望ましい。上記したＣＢＲ及びＶＢＲサービスは、いずれも、リアルタイムトラフィックを処理する。

【０００８】ＡＢＲ及びＵＢＲサービスは、ノンリアルタイムトラフィックを扱うサービスであり、主に、コンピュータ通信に用いられる。従来、ＡＢＲトラフィックは、閉ループフィードバックを用いて制御され、これは、約３％のオーバーヘッドを考慮している。

【０００９】一般に、ソースは、ネットワーク中を伝搬するリソース管理セル（ＲＭセル）を生成する。各ＲＭセルがスイッチを通過する際、サポート可能なレート、すなわち、ソースがデータを送出すべきレート（一般に明示レートと呼ばれる）を示すように、ＲＭセルは更新される。これらのＲＭセルはソースにフィードバックされ、ソースはその伝送レートをＲＭセルの明示レートに応じて調整することができる。但し、このようなフィードバックシステムはかなりの遅延を有し、したがって、リアルタイムトラフィックには用いることができない。

【００１０】一方、サービスクラスに応じて、ソースは適切なサービス品質（ＱｏＳ）を要求する。一般に、ＱｏＳは、伝送遅延と、セル損失率と、セル損失遅延変動に基づいて決定される。上述のように、呼を受け付けた場合でも、例えば、リーキバケットを用いてピークレートを制御することにより、ソースにおける伝送を調整することができる。したがって、コネクション設定において、ソースは、適切な使用量パラメータ制御（ＵＰＣ）値を交渉し、望ましいＱｏＳを申告する。次に、コネクション受付制御（ＣＡＣ）により、ネットワークが呼をサポートできるかどうかを決定する。

【００１１】ソースは、さらに、宛先アドレスも指示する。宛先アドレスを用いて、ＡＴＭネットワークは、仮想チャネル（ＶＣ）を設定し、適切なＶＣインディケータをソースに与える。ソースは、各伝送セル内にＶＣインディケータを挿入する。チャネルは、呼の継続中は、不変である。すなわち、呼の全てのセルは、同一チャネルを経由してルーティングされる。しかしながら、当該チャネルは他のソースと共有されることもあるため、仮想チャネルと呼ばれる。このことは、チャネルとソースの間には、一対一の対応関係がないことを意味している。

【００１２】一般に、受け付けた呼は、ＡＴＭスイッチ内のあるバッファに対応づけられ、スケジューリングアルゴリズムによって、いずれのバッファ、すなわち、いずれの呼が目的の時間に処理されるかを決定する。スケジューリングは、好ましくは、呼受付時に保証されたＱｏＳを考慮し、ネットワーク資源によって公平に共有されるべきである。また、このスケジューリングにおける

アルゴリズムは動作維持性を有していること、すなわち、セルがバッファ内に存在している場合には、停止すべきではないことが望ましいとされている。

【００１３】この種、ＡＴＭスイッチを含むＡＴＭ交換システムが、英国特許第２，２７２，８２０号（以下、引用例と呼ぶ）において提案されている。このＡＴＭ交換システムは、データトラフィックがピークの状態にある期間においても、ＡＴＭスイッチ動作を容易に行うことができるＡＴＭ交換システムが開示されている。この引用例に示されたシステムは、複数の入力ポート及び複数の出力ポートを備えたＡＴＭスイッチ、当該入力ポートに接続された入力ポートサーバ、及び前記出力ポートに接続された出力ポートサーバとを有しており、各入力ポートサーバには、複数のバッファが設けられている。このシステムでは、データを各バッファから出力ポートサーバに送信する前に、各バッファから通信すべき出力ポートサーバに帯域幅要求を出し、出力ポートサーバにおいて、処理可能な容量があるかどうかが決定されている。容量が或る場合には、ＡＴＭスイッチ動作が行われ、他方、容量がない場合には、キューイングが行われている。

【００１４】更に、ＡＴＭネットワークに使用できるスイッチとしても、種々のスイッチが提案されている。これら提案されたスイッチは、一段構成あるいは小さい一段スイッチを多段に接続した多段構成を有している。スイッチは、一般に、セルバッファの位置によって、入力バッファ型及び出力バッファ型かに分類できる。出力バッファを設けることによって最適なスループットが達成されることは良く知られている（例えば、Ｍ．Ｊ．Ｋａｒｏｌ、Ｍ．Ｇ．Ｈｌｕｃｈｙｊ、Ｓ．Ｐ．Ｍｏｒｇａｎによる「Ｉｎｐｕｔ ｖｓ．Ｏｕｔｐｕｔ Ｑｕｅｕｅｉｎｇ ｏｎ ａ Ｓｐａｃｅ－Ｄｉｖｉｓｉｏｎ Ｐａｃｋｅｔ Ｓｗｉｔｃｈ」（ＩＥＥＥ Ｔｒａｎｓ．Ｃｏｍｍ．，Ｖｏｌ．３５、１３４７～１３５６頁、１９８７年１２月）参照）。

【００１５】

【発明が解決しようとする課題】上記した引用例に示されたＡＴＭ交換ネットワークでは、出力ポートサーバに、帯域幅を受付けられる容量があるか否かが決定されている。より具体的に言えば、入力ポートに負荷が増えてきた場合、空き帯域があるかどうかを問い合わせ、ある場合には、帯域を予約しておく。逆に、負荷が減ってきた場合、予約している帯域を開放しなければ、他のクラス／ポートがその帯域を利用できないため、その帯域が無駄になってしまう。また、実際、負荷の変化は、非常に激しく、且つ、予想しにくいため、引用例に示されたネットワークにおいては、トラフィックの変化に追随できず、このため、帯域の利用率が悪くなってしまう。

【００１６】更に、帯域幅要求があった場合に、帯域幅自体を変化させることについて、引用例では、何等考慮されていない。更に、この引用例においては、複数のサービスクラスに適用することについても、全く開示していない。

【００１７】一方、出力バッファを有するスイッチ構成では、出力バッファは、ラインレートのＮ倍のアクセス速度で動作することが必要である。ここで、Ｎは入力ポートの数である。スピードアップ係数Ｎは、いわゆる「ノックアウト原理」（Ｙ．Ｓ．Ｙｅｈ、Ｍ．Ｇ．Ｈｌｕｃｈｙｊ、Ａ．Ｓ．Ａｃａｍｐｏｒａによる「Ｔｈｅ Ｋｎｏｃｋｏｕｔ Ｓｗｉｔｃｈ：Ａ Ｓｉｍｐｌｅ，ＭｏｄｕｌａｒＡｒｃｈｉｔｅｃｔｕｒｅ ｆｏｒ Ｈｉｇｈ－Ｐｅｒｆｏｒｍａｎｃｅ Ｐａｃｋｅｔ Ｓｗｉｔｃｈｉｎｇ」（ＩＥＥＥ Ｊ．Ｓｅｌｅｃｔ．Ａｒｅａｓ Ｃｏｍｍ．Ｖｏｌ．５、１２７４～１２８３頁、１９８７年１０月）参照）を用いることにより、Ｎ＝８まで下げることができる。しかしながら、スイッチに不均一なトラフィックパターンが加えられた場合には、望ましくないセル損失が生じることがある。共有メモリを有するスイッチにおいても、共有メモリはＮ倍アクセススピードを持たなければならない。

【００１８】入力バッファを備えたスイッチはスピードアップを一切必要としないが、ヘッドオブライン（ＨＯＬ）ブロッキングによるスループット低下が生じる。すなわち、宛先出力回線が先頭ラインセルを受け付ける準備ができるまで、入力バッファキューの先頭のセルがバッファ内の他の全てのセルをブロックする。しかしながら、先頭ラインセルによりブロックされている他のセルを他の宛先出力回線では実際には受け付けることができる場合もある。このように、受付可能なセルを不所望にブロックすることは帯域幅の効率的な使用を妨げ、不必要な遅延を引き起こすことになる。

【００１９】現在のＡＴＭスイッチは比較的単純なスケジューリングと、ＱｏＳに対するサポートが限定されたバッファ管理メカニズムを有する。一方、ＡＴＭ技術が広域網（ＷＡＮ）市場に普及するにつれて、より多様なアプリケーションから、かなり大量のトラフィックを扱うＷＡＮコアでは、より高度なスイッチが必要となる。次世代ＷＡＮコアスイッチは、大容量と、マルチクラスのトラフィックに対してＱｏＳサポートを行なえる能力を備えることが必要であると考えられる。

【００２０】したがって、本発明の目的は、このような多様なトラフィックをサポートできるＡＴＭスイッチを提供することである。

【００２１】本発明の他の目的は、帯域の変化に容易に対応出来るＡＴＭスイッチを提供することである。

【００２２】本発明の他の目的は、帯域幅を変化させることができるＡＴＭスイッチを提供することである。

【００２３】

【課題を解決するための手段】本発明によれば、複数のトラフィッククラスとサービス品質（ＱｏＳ）保証をサ

ポートする大容量ＡＴＭスイッチが得られる。このスイッチは、ＣＢＲやＶＢＲなどのきびしいＱｏＳ要求をともなうリアルタイムトラフィッククラスと、ＡＢＲやＵＢＲなどのそれほどきびしくない要求をともなうノンリアルタイムトラフィッククラスの両方をサポートする。この構成は、さらに、リアルタイム及びノンリアルタイムのマルチキャストフローを効率的に調整できる。スイッチは、大容量バッファを備えた入力／出力モジュールを相互接続する高速コアスイッチモジュールを備えると共に、入力及び出力バッファの双方を有している。この場合、コアスイッチモジュールには、小容量の高速のコアバッファが設けられる。このように、入力バッファ、コアバッファ、及び、出力バッファの３バッファ構成が採用され、バックプレッシャー、即ち、出力バッファから入力バッファへのフィードバックを掛けることにより、入力ポート間の制御を避けることができ、且つ、大容量の出力バッファの速度を低下させることができる。尚、入力バッファは、出力ポートに対応して設置されても良いし、出力ポート並びに出力回線に対応して設置されても良い。入力バッファを出力ポートに対応して設けた場合、出力ポートへのトラヒックをスケジューラで制御できるが、出力回線の区別がないため、大きな出力バッファが必要となる。一方、入力バッファを出力ポート並びに出力回線に対応して設けた場合、出力回線へのトラヒックを制御できるため、出力バッファの容量を減少させることができる。

【００２４】このように、インテリジェントスケジューリング及びキュー管理メカニズムによって、コアスイッチモジュールは制御されたクラスにしたがってアクセスされる。

【００２５】スイッチは、好ましくは、本発明者らにより発明され、同日付で提出される新規なスケジューリング方法、即ち、動的レート制御（ＤＲＣ）とともに用いられる。このＤＲＣは、内部輻輳を制御し、スイッチのボトルネック部で、競合するフロー間の公平なスループット性能を達成する。これは、各ボトルネック部において比例微分[proportional-derivative]（ＰＤ）制御器を用いた閉ループ制御を行なうことにより達成される。ＤＲＣ方式は最低保証レートに、未使用帯域幅を公平に配分する動的レート成分を加えたレートを、各フローに保証する。このことは、異なるトラフィックサービスクラスに対するＱｏＳを統合したシステムにおける基盤を与えることができる。

【００２６】本発明に係わる大容量スイッチでは、ＤＲＣスケジューリングメカニズムは、インテリジェントキュー管理メカニズムとともに動作する。ＤＲＣスケジューラは、スイッチ内のボトルネック部において輻輳を検出し、スイッチの入力側に、セルを移動させることにより、輻輳を緩和する。ここで、初期パケット廃棄（ＥＰＤ）や部分パケット廃棄（ＰＰＤ）などのセル廃棄メカニズムを、個々のクラスキューに適用してもよい。また、低い優先順位をタグ付けされたセル、すなわち、セル損失優先（ＣＬＰ）ビットが１に設定されたセルは、キューがしきい値を越えると、廃棄される。ＤＲＣを用いると、スイッチ輻輳レベルにしたがって制御され、セルが廃棄されるため、セル廃棄メカニズムはより効果的に動作する。

【００２７】ここに述べる本発明の大容量スイッチは、総スループット及びマルチクラスＱｏＳに対するサポートの両方において、現在のスイッチに対して著しく改善されている。また、本発明に係わる大容量スイッチはフレキシビリティとスケラビリティを有しており、現在及び将来の高性能ＡＴＭネットワークの要望を満足させることができる。

【００２８】

【発明の実施の形態】

１．全般的な構成

好ましい実施例において、本発明の大容量スイッチは一段スイッチであり、この一段スイッチは、より大容量の多段スイッチにおけるスイッチング素子として用いることもできる。本発明による大容量スイッチは、入出力バッファを備えたスイッチ（以下、入出力バッファ付スイッチと呼ぶ）（Ｒ．Ｆａｎ、Ｈ．Ｓｕｚｕｋｉ、Ｋ．Ｙａｍａｄａ、Ｎ．Ｍａｔｓｕｕｒａによる「Ｅｘｐｅｎｄａｂｌｅ ＡＴＯＭ Ｓｗｉｔｃｈ Ａｒｃｈｉｔｅｃｔｕｒｅ （ＸＡＴＯＭ） ｆｏｒ ＡＴＭ ＬＡＮｓ」（Ｐｒｏｃ．ＩＣＣ´９４、９９～１０３頁、１９９４年５月）参照）として分類できる。入出力バッファ付構造の目標は、入力バッファ及び出力バッファの長所を組み合わせることである。

【００２９】好ましい実施例において、出力バッファは、コアスイッチモジュールの一部である小型で高速のバッファである。セルはライン速度で動作するバッファを有する入力モジュール内で主にキューイングされる。宛先出力ポート、あるいは、宛先出力回線にしたがって、入力モジュール内のセルをキューイングすることによりＨＯＬ（ヘッド・オブ・ライン）ブロッキングは回避される。この構成は、高速で大容量の出力バッファを必要とすることなく、出力バッファ付きを用いた場合と同等のスループットを達成できる。その上、入力ポートにバッファを設けることは、出力バッファを設ける場合よりも有効である。セル損失性能を同じにした場合には、出力ポートにおいてキューイングするよりも入力ポートにおいてセルのキューイングを行った方が、全体としてはバッファの容量を少なくできる。

【００３０】本発明は、高速且つ簡単であり、クラスに関係なく、しかも、無損失でスイッチングを行なえる大容量マルチクラススイッチを提供することを主目的としており、このスイッチは、新規な高速コアスイッチ素子によって構成されている。コアスイッチモジュールに対

するアクセスは、入力モジュール（ＩＭ）のインテリジェントスケジューリングメカニズムにより制御される。入力モジュールは、クラス毎のキューイング、或いは仮想チャネル毎のキューイングができるように構成されている。

【００３１】大容量スイッチの好ましい実施例の全体図を図１に示す。図示の実施例において、コアスイッチモジュール１０は、１６個の入力ポートＩＰ１～ＩＰ１６と、１６個のユニキャスト出力ポートＯＰ１～ＯＰ１６と、１個のマルチキャスト出力ポートＭＯＰによって構成され、これら全ては、高速ＴＤＭバス２０に接続されている。図示の実施例において、コアモジュールの入出力ポートは２．４Ｇｂｐｓのレートで動作し、ＴＤＭバスは４０Ｇｂｐｓのレートで動作する。この例では、１セル時間内に、コアスイッチモジュールは、各入力ポートからの一つのセルを出力ポートのいずれかに交換することができるものとする。

【００３２】入力モジュール（ＩＭｉ）は、コアモジュールの各入力ポートＩＰｉに接続される。ここで、入力モジュールＩＭｉの出力回線容量は２．４Ｇｂｐｓであるものとする。入力モジュールＩＭｉの入力側は入力回線ＩＬ１ｉ～ＩＬ１６に接続され、これらの入力回線は次の３通りのいずれかの構成を有している。すなわち、１）１本の２．４Ｇｂｐｓ回線、２）４本の６２２Ｍｂｐｓ回線、３）１６本の１５５Ｍｂｐｓ回線のいずれかによって構成される。図示された実施例のすべての場合においても、入力モジュールＩＭｉの総入力回線容量は２．４Ｇｂｐｓである。入力回線ＩＬ１ｉ～ＩＬ１６ｉには、異なるＱｏＳ要求をともなう異なるクラスに分類されるソースから、伝送が行なわれる。

【００３３】仮想チャネル（ＶＣ）毎のキューイング（図２ａ）と、出力ポートに対するクラス毎のキューイング（図２ｂ）と、出力回線に対するクラス毎のキューイング（図２ｃ）に応じて、入力モジュールは構成される。ＶＣ毎のキューイングは、スループットの点では最良であるが、各入力モジュールに多数のバッファを必要とする。したがって、ＶＣ毎のキューイングは設計上では望ましいが、実現性の面では好ましくない。

【００３４】図２ｂに示すように、出力ポートに応じたクラス毎のキューイングにおいて、各入力モジュールＩＭｉは、出力モジュールの数に対応して、多数のレイヤ／プレーンＬＯＰ１～ＬＯＰ１６、本実施例では、１６個のレイヤを有している。各レイヤは、サポートする必要のあるクラス数ｋに対応して数個のバッファＩＢ１～ＩＢｋを備えている。また、各入力バッファＩＢｉには、サービスレートＲｉ１～Ｒｉｋが保証されている。これにより、入力セルは、宛先出力ポートに対応する適正なレイヤに、そして、そのクラスに応じたレイヤ内の適切な入力バッファにルーティングされる。出力ポート上の負荷に対応して、ＤＲＣレートが出力モジュールからフィードバックされる。

【００３５】図２ｃに示すように、出力回線に応じて、クラス毎のキューイングを行なうためにスイッチが構成されている場合、入力モジュール内には２組のレイヤ／プレーンが設けられている。まず、出力モジュールは出力ポートに対応する出力ポートプレーンＯＰ１～ＯＰ１６Ｉに分割されている。次に、各出力ポートプレーンを出力回線に対応する出力回線プレーンＯＬ１～ＯＬｋに分割されている。最後に、各出力回線プレーンは各クラスに対応する複数個のバッファを備えている。この場合、２個のＤＲＣレートがフィードバックされ、一つは、出力ポート上の負荷、他方は、出力回線上の負荷をそれぞれ示している。

【００３６】同様に、出力モジュール（ＯＭｉ）を各ユニキャスト出力ポートＯＰｉに接続する。ＯＭの出力側は、２．４Ｇｂｐｓの総出力回線容量を有する１本、４本、あるいは、１６本の出力回線からなる。出力モジュールは、出力回線に対応する出力プレーンに分割される。各出力プレーンは、サポート可能なクラスに対応する複数個のバッファを有する。

【００３７】コアスイッチモジュール１０の各出力ポート（ＯＰｉ）は、２個の小さい出力バッファＲＴｉ、ＮＲＴｉに組み合わされ、一方のＲＴｉは、リアルタイムトラフィック用に指定され、他方のＮＲＴｉはノンリアルタイムトラフィック用に指定されている。好ましい実施例において、各出力バッファＲＴｉあるいはＮＲＴｉは、２００個程度のセルを記憶することができる。各ユニキャスト出力ポートの出力において、ノンリアルタイム及びリアルタイムバッファ用出力回線は、マルチキャスト出力ポートの対応する出力回線と組み合わされる。各セル時間のあいだ、単一のセルがユニキャスト出力ポートＯＰｉの出力から、対応する出力モジュールＯＭｉに伝送される。

【００３８】図１を参照すると、優先順位は以下の１）、２）、３）、４）の順で低くなる。１）マルチキャストリアルタイムトラフィック、２）ユニキャストリアルタイムトラフィック、３）マルチキャストノンリアルタイムトラフィック、４）ユニキャストノンリアルタイムトラフィック。

【００３９】コアスイッチモジュールがユニキャスト及びマルチキャスト交換を実行するための基本ハードウェアを提供するが、好ましい実施例において、当該スイッチの本発明の対象となる部分は、実質上、入力モジュール（ＩＭｉ）と出力モジュール（ＯＭｉ）にある。各入力／出力モジュールには、スケジューラと、キュー管理部、即ち、マネージャーと、大容量のスペースを有するバッファが設けられている。コネクション受付制御器と入力／出力モジュール間とは、密に結合されているため、キューフローがそのサービス品質（ＱｏＳ）要求を確実に満足させることができる。

【００４０】各入力モジュールＩＭは、ＡＴＭセルヘッダ変換動作を行なうと共に、ＶＣ（図２ａ）及びクラスとコアスイッチ宛先出力ポート（ＯＰｉ）（図２ｂ）、あるいは、クラスと宛先出力回線（図２ｃ）により構成されるキュー内の入力セルが格納される。ここで、キューフローの用語は、与えられたキューに対応する全てのコネクションの総トラフィックをあらわすために用いられる。コネクション受付制御器（ＣＡＣ）（図３参照）はどのトラフィッククラスのキューに対しても、フレキシブルにＱｏＳを割り当てることができ、ここでは、このＱｏＳをプログラマブルＱｏＳと呼ぶ。

【００４１】一方、クラスキューは、リアルタイムあるいはノンリアルタイムキューに分類される。

【００４２】各セル時間中、入力モジュールＩＭｉ内にキューがあれば、スケジューラがキューのうちの一つを選択する。選択されたキューから、ヘッドオブライン（ＨＯＬ）セルがＴＤＭバスを経由して宛先出力ポートＯＰｉに伝送される。キューマネージャーはキューにセルバッファを割り当て、バッファしきい値を超過した場合には、セルを廃棄する。入力モジュールＩＭｉ内のセルのキューイングは、コアスイッチモジュール内の出力ポートにおける輻輳を回避するように設計されている。異なる入力モジュールＩＭｉからのキューフローの合計が出力ポートＯＰｉにおける容量Ｃを越える場合に、輻輳が生じることもある。このような状況のもとで、出力ポートＯＰはボトルネック部となる。

【００４３】出力モジュールの構成は、入力モジュール構成と同様である。各出力モジュールは、基本的にスイッチの残りの部分から独立して、且つ、判断機能を備えたインテリジェント多重分離器として動作する。出力モジュールＯＭｉ内で、セルはクラス及び宛先出力回線にしたがってキューイングされる。出力モジュールＯＭｉにおけるセルのキューイングは、出力モジュールＯＭｉに接続された出力回線ＯＬｉにおける輻輳によって生じる。すなわち、出力回線に対するキューフローの合計がその容量を越えた場合に、出力モジュールにおけるセルキューが生じる。このように、出力回線ＯＬｉは、レートの不適合が生じた場合、スイッチ内のキューフローに対する他のボトルネック部となる。出力ポートＯＰｉ及び出力回線ＯＬｉのようなボトルネック部における内部スイッチ輻輳は、入力／出力モジュール内のインテリジェントスケジューリングとキュー管理によって制御される。

【００４４】図３は、入力及び出力モジュールの構成をより詳細に示している。図示された入力モジュール３０は、１６個の出力ポートＯＰ１～ＯＰ１６に対応する１６個のプレーンを備えている。各プレーンは複数個の（４個のみ図示されているが、より多くてもよい）同一構成のバッファ３２を備えている。これらのバッファ３２はＣＡＣ３３により、それぞれのＱｏＳ要求を有する多様なクラスに対応するようにプログラムされている。図示された入力モジュール３０において、４個のバッファはＣＢＲ、ＶＢＲ、ＡＢＲ、ＵＢＲの４つのサービスクラスにそれぞれに割り当てられる。ＣＡＣ３３は、さらに、バッファ内のセルに対して、送受信時点を指示するタイムスタンプを与える。

【００４５】コアスイッチモジュール３４は、複数個のバッファ付き出力ポートｉに接続されたＴＤＭバス３５を備えている。図示の例では、各出力ポートは２個のバッファ、すなわち、リアルタイムバッファＲｔとノンリアルタイムバッファＮｒｔを有し、出力バッファは出力モジュールに接続されている。各出力モジュールは出力回線に対応するプレーンＯＬ１～ＯＬｋに分割されている。さらに、これらの出力プレーンには、複数個のプログラマブルバッファ及び１個のスケジューラが備えられている。

【００４６】２．マルチキャスティング

本発明によるスイッチの特徴は、ユニキャストだけでなくマルチキャストを効率的にサポートできる点にある。スイッチ内のユニキャストコネクションは、入力モジュールＩＭｉへの入力回線から、出力モジュールＯＭｉの単一の出力回線までのコネクションである。一方、マルチキャストコネクションは、入力モジュールＩＭｉから、一つ以上の出力モジュールＯＭｉにおける複数の出力宛先回線との間のコネクションである。コアスイッチモジュール３４は、複数の出力モジュールＯＭｉ間のマルチキャストをサポートできる。２つ以上の出力モジュールＯＭｉにマルチキャストされるセルは、入力モジュールＩＭｉからマルチキャスト出力ポートＭＯＰに送られる。

【００４７】当該マルチキャストセルも、リアルタイムあるいはノンリアルタイムセルに区分され、当該セルのコピーが一つだけマルチキャスト出力ポートＭＯＰ内の対応するバッファに格納される。出力モジュールＯＭｉに送出される直前に、マルチキャスト出力ポートＭＯＰの出力側でセルが複写される。

【００４８】図１に関連して述べられたように、リアルタイムマルチキャストトラフィックは各出力モジュールＯＭｉに入力される場合、最優先順位を有している（図１において、出力における優先順位は、出力モジュールに対する入力をあらわし、且つ、垂直線方向に向いた矢印の順番により示されている。したがって、垂直線に対して一番上にある矢印が最優先順位を示している。）。

【００４９】与えられたセル時間内に、マルチキャスト出力ポート用リアルタイムバッファの先頭にマルチキャストセルがある場合、セルは複写され、マルチキャストに関与する出力モジュールＯＭｉに送られる。ここで、マルチキャストセルの複写は、コアスイッチモジュールのＴＤＭバス上では行われない。

【００５０】リアルタイムマルチキャストトラフィック

は、最優先順位を有するため、ブロッキングされることはない。一方、ノンリアルタイムマルチキャストトラフィックは、リアルタイムマルチキャスト及びリアルタイムユニキャストよりも低優先順位を有している。

【００５１】このことを考慮すると、与えられたセル時間内において、出力ポートＯＰｉにリアルタイムマルチキャストセルが或る場合、或いは、マルチキャストの対象となるユニキャスト出力ポートＯＰｉのいずれかにユニキャストリアルタイムセルがある場合には、出力ポートＯＰｉにおけるノンリアルタイムマルチキャストセルはブロックされることになる。

【００５２】３．フィードバック制御

フィードバック制御は、スイッチの効率的な動作を保証するために用いられる。コアスイッチモジュール内の出力ポートバッファは小規模であるため、すぐにオーバーフローする可能性がある。そこで、このようなオーバーフローを調整するために、２つの基本的なフィードバックメカニズムが用いられる。これらのフィードバックメカニズムには以下のようなメカニズムがある。

【００５３】１．　出力ポートバッファにおけるキューを短く保ちながら、ボトルネックレートに合致させ、且つ、利用度を高く維持した閉ループフィードバックメカニズム（第一の制御メカニズム）。

【００５４】２．　コアスイッチモジュールの出力ポートバッファが、第一の制御メカニズムによる制御にもかかわらずオーバーフローする可能性を有する場合に起動されるしきい値ベースのレートフィードバックメカニズム（第二の制御メカニズム）。

【００５５】第一の制御は、入力モジュールにおける動的レート制御（ＤＲＣ）スケジューリングにより達成される。第二の制御は、コアスイッチモジュール内に組み込まれ、出力ポートＯＰｉボトルネックにおける短期的輻輳を迅速に制御するための安全メカニズムとみなされる。コアスイッチモジュールは、各セル時間中、全ての入力モジュールＩＭｉに出力ポートＯＰｉの状態情報を同報的に送るためのフィードバック経路を有している。フィードバック信号が出力ポートから入力モジュールに伝搬する時間は、本明細書中にτｄ（セル時間）であらわされ、この量はシステムに応じて異なる。

【００５６】好ましい実施例において、スケジューラは、入力モジュールに対して未使用帯域幅を配分する。この場合、入力モジュールからの実際の伝送レートは、保証最小レートを越えることもある。しかしながら、状況によっては、全ての利用可能な帯域幅を用いた場合、ある出力ポートで輻輳を生じることがある。そこで、好ましい実施例において、このような輻輳を少なくするためにフィードバック信号が使用される。

【００５７】好ましくは、入力ポートバッファに関連して、３つのしきい値が、制御フィードバック信号として生成される。これらしきい値は、１）ストップリアルタイム（ＲＴ）、２）シェーピングリアルタイム（ＲＴ）、３）ストップリアルタイム（ＮＲＴ）がある（図１参照）。ストップＲＴしきい値インディケータは、リアルタイムバッファにおける格納量がしきい値Ｔｈstop以上である場合、１に設定され、それ以外の場合には、ストップＲＴインディケータはゼロとなる。同様に、ノンリアルタイムキューの格納量がＴｈstop以上である場合、ストップＮＲＴは１となる。ストップ信号が入力モジュールＩＭに到達するまで、全ての入力モジュールＩＭが同一の出力ポートにセルを送出するという最悪の状態でバッファオーバーフローが生じないように、ストップしきい値Ｔｈstopが最大値として選択されている。シェーピングＲＴインディケータは、リアルタイムバッファがシェーピング用しきい値Ｔｈshape （＜Ｔｈstop）以上の場合、１に設定される。表１に、しきい値インディケータと２ビットに符号化された出力ポートに対する制御信号Ｂ０、Ｂ１との関係、入力モジュールにより実行される動作が示されている。

【００５８】

【表１】

フィードバック制御信号

しきい値インディケータ			制御ビット		入力モジュールで実行される動作
シェーピングRT	ストップRT	ストップNRT	Ｂ１	Ｂ０	
0	0	0	0	0	送信RT、停止NRT
0	0	1	0	1	送信RT、停止NRT
1	0	0	1	0	シェーピングRT、送信NRT
1	1	1	1	1	停止RT、停止NRT

ストップＲＴインディケータが１に設定されると、適切なフィードバック信号が送出される。τｄセル時間の後、信号は全ての入力モジュールに与えられ、各入力モジュールは、対応する出力ポートにセルフロー（リアルタイム及びノンリアルタイムの両方）を制限する。ストップＮＲＴインディケータは、ノンリアルタイムトラフィックに対して、同様に機能する。上記したフィードバック信号により、出力ポートにおけるセル損失を回避することができる。ここで、ストップ信号は、入力モジュール内の入力のキューイングを発生させる。ストップ信号がない場合、入力モジュール内におけるセルのキューイングは生ぜず、結果として、出力ポートにおいてオーバーフローが生じて、セル損失が発生してしまう。

【００５９】シェーピングＲＴインディケータは、キュ

ーフローに対して予め割り当てられた最低保証レートに基づき、リアルタイムトラフィックの輻輳を制御する手段として動作する。

【００６０】ここで、ある出力ポートＯＰｊから入力モジュールＩＭｉに対してシェーピングＲＴ信号が受信されると、出力ポートＯＰｊに対応する全てのリアルタイムキューフローは、その最低保証レートにシェーピングされる。すなわち、リアルタイムキューは、利用できる未使用帯域幅に関係なく、その最低レートでスケジューリングされる。この動作により、リアルタイムキューフローに対する最低保証スループットを確保しながら、出力ポートＯＰｊにおけるリアルタイムキューの格納量が大きくなるのを防ぐことができる。

【００６１】このように、リアルタイムキューフローは、輻輳がない場合、最低保証レートより大きいレートで処理され、輻輳がある場合には、最低保証レートに等しいレートで処理される。

【００６２】以下に詳述するように、ＤＲＣ（動的レートコントロール）スケジューラでは、リアルタイムキューフローに対しては、ストップ信号が低い確率しか発生しない。

【００６３】４．トラフィッククラス及びサービス品質（ＱｏＳ）サポート

４．１　リアルタイムトラフィック

ＣＢＲやＶＢＲなどのリアルタイムトラフィックでは、セル遅延と、セル損失と、セル遅延変動（ＣＤＶ）に対して厳しい要求が加えられている。コネクション受付制御（ＣＡＣ）アルゴリズムについて言えば、大容量スイッチ構造を採用することによりＱｏＳ保証を提供することができる。このようなＣＡＣアルゴリズムとして、Ｇ．ＲａｍａｍｕｒｔｈｙとＱ．Ｒｅｎによる「Ｍｕｌｔｉ－Ｃｌａｓｓ Ｃｏｎｎｅｃｔｉｏｎ Ａｄｍｉｓｓｉｏｎ Ｃｏｎｔｒｏｌ Ｐｏｌｉｃｙ ｆｏｒ Ｈｉｇｈ Ｓｐｅｅｄ ＡＴＭ Ｓｗｉｔｃｈｅｓ」（ｐｒｏｃ．　ＩＥＥＥ　ＩＮＦＯＣＯＭ　'９７（神戸、日本）１９９７年４月）によって提案された手法を本発明に係わるスイッチのＣＡＣにおいても用いることができる。このＣＡＣは、リアルタイムキューフロー内のコネクション全てのＱｏＳ要求を満足させるように、必要な帯域幅を算出する。すなわち、ＣＡＣは、任意のキューフロー内のコネクションの統計的多重化を考慮している（統計的多重化は、全ストリームに必要な帯域幅が各ストリームに必要な個々の帯域幅の合計よりも少ないということを考慮した手法である）。必要な帯域幅は、各コネクションに対するＵＰＣパラメータ及びフローに対するバッファの予備割当て量に基づいて算出される。しかしながら、この従来技術では、計算された最低レートはＣＡＣの目的のためだけに用いられ、スケジューラでは使用されていない。

【００６４】ＤＲＣスケジューリングメカニズムは、各キューフローがその最低保証レートを受けとり、したがって、ＱｏＳがフロー内のコネクション全てについて確実に保証する。コアスイッチ素子においてリアルタイムトラフィックがノンリアルタイムトラフィックに対して高い優先権を持ち、出力ポートからのシェーピングフィードバックメカニズムにより、輻輳状態においても、当該キューフローの最低保証レートが保証されるから、最低レートは短時間内においても保証される。すなわち、輻輳状態のもとでは、シェーピングフィードバックメカニズムにより、未使用帯域幅の分配が停止され、これにより、レートを最低保証レートに低下させて輻輳を緩和する。この状態では、最低保証レートが確実に保証される。さらに、シェーピングモードにおいて、最低レートでの動作を行なっているキューでは、サービスを受けることができる場合、当該優先ビットがセットされる。

４．２　ノンリアルタイムトラフィック

ノンリアルタイムトラフィッククラスにはＡＢＲ及びＵＢＲが属している。これらのクラスは、通常、ＱｏＳに対する要求は厳しくないが、最低スループットが要求されることがある。ノンリアルタイムキューフローについての最低レートは、フロー内のコネクション全てについての最低スループットの合計に等しい。大容量スイッチスケジューラは、ＤＲＣスケジューリングを行なうことによって、各ノンリアルタイムフローに対して最低レートを保証できる。スイッチボトルネックにおける未使用帯域幅も、競合するキューフロー（リアルタイム及びノンリアルタイムの両方）間で分配される。未使用帯域幅の分配は、異なるトラフィッククラスに割り当てられた重みω１に依存して行なわれ、好ましくは、これらのレートは動的に割り当てられる。

【００６５】ＵＢＲソースはレート制御されず、従来のスイッチ構造においてスループットの損失を引き起こし得る。動的レート制御では、ＵＢＲキューには、その最低レートに、未使用帯域幅の公平な割当を加えたものが与えられる。ＡＢＲソースは閉ループフィードバックメカニズムによってレート制御される。スイッチにおいて、明示レート（ＥＲ）値が、コネクションフローにおける各ボトルネック部において算出される。大容量スイッチ構造において、ＡＢＲ　ＥＲ値は、出力モジュール内の出力ポートボトルネック部及び出力回線ボトルネック部において算出される。ＥＲを算出するために様々な方法が用いられる。しかしながら、好ましい実施例において、ＡＢＲ　ＥＲ値はＤＲＣレートの算出と同様に算出される。

【００６６】５．動的レート制御

動的レート制御（ＤＲＣ）は、大容量スイッチにおいて用いられるセルスケジューリングのためのメカニズムである。ＤＲＣについては米国特許出願番号０８／９２４，８２０号に記載されているが、説明の便宜上、本発明によるスイッチに適用した場合の、ＤＲＣの基本原理

の概要を以下に説明する。また、A．KolarovとG．Ramamurthyによる「Design of a Closed Loop Feed Back Control for ABRService」（proc．IEEE INFOCOM ´97（神戸、日本）１９９７年４月）には、ATMネットワークに適用した場合のABRサービスに対するフィードバック制御が記載されている。しかしながら、以下の説明におけるフィードバックは、ATMスイッチに適用されていることは留意すべきである。

【００６７】DRCの基本原理は、各クラスのキューを仮想ソースと同様に扱い、そのサービスレートを動的に調整することにより、スイッチ内のボトルネック部における利用可能な未使用帯域幅に反映させることである。すなわち、各クラスは、その最低保証レートに、利用可能な未使用帯域幅の動的に調整され、公平に配分されたレートを加えたレートで処理される。スケジューリングは、キューサービスレートを算出し、全てのキューについてレートシェーピングを行なうことによって実現される。スケジューリングに対するこの方法の重要な特徴は、全てのキューを平均的なキューの集合に変え、クラスのQoSを、そのクラスに対して保証された帯域幅によって決定することである。ここで、平均的なキューは、CACにより各クラスに割り当てられる。

５．１　最低保証レート

スイッチにおける任意のコネクションに対するQoSを保証するために、スイッチでは、トラフィック特性と帯域幅リソース間のマッピングが行なわれる。任意のコネクションiのトラフィック仕様には、セル損失確率と、遅延、および／または遅延ジッタであらわされた一組のQoS要求が含まれている。指定された要求の全てに関連したアルゴリズムを実現するよりも、これらの要求全てを関連した単一の変数にこれらの要求をマッピング、即ち、変換するほうが簡単である。

【００６８】DRCの好ましい実施例では、要求が帯域幅、又は、レートMiにマッピングされている。具体的に言えば、コネクションiにレートMiを割り当てることにより、当該QoS要求が満たされることになる。好ましい例において、Miは、コネクション受付制御（CAC）アルゴリズムによって与えられる。この場合、レートMiは、QoSに対する要求の全てを満たすように、十分に近似されなければならない。

【００６９】一旦、Miが決定されると、スケジューラはコネクションiに対して最低レートを保証する。このことは、コネクションiのQoSが保証されることを意味している。このように、上記したスケジューラはひとつの変数のみを考慮すれば良いから、構成によって簡単になる。

【００７０】ここで、N個のコネクションが容量Cのリンク上に、多重化されるネットワーク内における集線位置について考察してみよう。この場合、次の数１式の関係が成立することは明らかである。

【００７１】

【数１】

$$\sum_{i=1}^{N} M_i \leq C$$

単純な先入れ先出し（FIFO）スケジューラを使用した場合、各コネクションに対して割当帯域幅Miを保証することはできない。例えば、あるコネクションがその割当分Miより高いレートで伝送し、これにより、別のコネクションから帯域幅を奪ってしまうことがある。

【００７２】コネクションがその割当分より多くの帯域幅を使用しないように保証する単純な方法は、各コネクションiのピークレートをMiに限定することである。このことは、例えば、リーキバケットなどの公知の方法を用いて、割り当てられた最低レートに各コネクションのピークレートをシェーピングすることにより実施することができる。

【００７３】図４は、各キューiを含むN個のキューがそれぞれのレートMi（i＝１，…，N）にシェーピングされた場合を示している。この例では、シェーピングされたトラフィックストリームは、次に多重化され、下流側バッファのレートC以下のレートでFIFO順に処理されている。

【００７４】ピークレートによる制限によって、全てのコネクションについて最低レートを保証することができる。しかしながら、このスケジューリング規則では、コネクションiは、広い帯域幅が利用可能な場合でさえも、割り当てられたMiより多くの帯域幅を用いることはできない。例えば、コネクションiがリンク帯域幅Cを共用する唯一のアクティブなコネクションである場合、全リンク容量Cが利用可能であっても、帯域幅Miだけしか使用できない。さらに、統計的多重化が行なわれることを前提として最低レートMiを算出する場合、リンク容量を超える確率は小さいため、全帯域幅が共有された場合を仮定して、コネクションiのQoSが保証される。この種の共用は、コネクションのピークレートが値Miに制限される場合には起り得ない。統計的多重化利得を前提として算出される割り当て帯域幅Miは、統計的多重化が行われない場合には、QoSを保証するには不十分になるおそれがある。

【００７５】一方、静的レートに基づくスケジューラのレート制御メカニズムは、開ループと呼ばれることがある。これは、スケジューラが動作維持型でないこと、すなわち、システム内で処理すべきセルがある場合でさえも、出力リンク上にセルが送出されない空白の時間があることである。この場合、利用可能な帯域幅がこれらのスケジューリング規則では、無駄になる可能性がある。

【００７６】他方、上記したDRCスケジューラは、閉

ループ制御メカニズムによってこの問題を解決している。

【００７７】ＤＲＣスケジューラの基本原理を図５に示す。前述のように、各トラフィックストリームは、リンクレートＣの共通のＦＩＦＯキューに入力される前に、ピークレートについてシェーピングされる。しかしながら、シェーピングレートＲｉは、リンク上で利用可能な帯域幅の量を反映するように動的に算出される。具体的に言えば、コネクションｉは、Ｒｉにピークレートシェーピングされるものとすると、Ｒｉは次式で与えられる。

【００７８】Ｒｉ＝Ｍｉ＋ｗｉＥ

ここで、Ｅは、ボトルネック部における推定未使用帯域幅（本明細書中では超過レートあるいは超過帯域幅とも呼ぶ）である。一方、ｗｉ≧０であり、静的あるいは動的に割り当てられる任意の重み係数である。通常、Ｅ≧０であるため、Ｒｉ≧Ｍｉが成り立つ。上式からも分かるように、コネクションｉには最低レートＭｉが保証されるが、未使用帯域幅が利用可能である場合には、より高いレートで伝送することもできる。逆に、輻輳中には、スケジューラがＥをゼロとすることにより、輻輳が解消されるまで、その最低保証レートでのみキューは処理されることがわかる。

５．２　閉ループレート制御

図６は、ＤＲＣスケジューラにおいてコネクションに分配される利用可能な帯域幅を検出する閉ループレート制御システムを示している。ここで、時間を、長さＴの区間に分割して、離散的にあらわすものとする。Ｘｉ（ｎ）はｎ番目の区間内に、コネクションｉにより発生されるセル数を示す。また、量Ｑ（ｎ）は第２段バッファ内のセル数を表わし、第１段バッファでは、各コネクションストリームは、次式で与えられるレートにしたがってシェーピングされるものとする。

【００７９】

Ｒｉ（ｎ）＝ｍｉｎ（Ｍｉ＋ｗｉＥ（ｎ），Ｃ）

コントローラ、即ち、制御器は、Ｑ（ｎ）が目標のキューしきい値Ｑ０に近い値に保たれるように、Ｅ（ｎ）を算出する。平衡状態において、十分なフローがソースから与えられている場合、第２段バッファへの全フローレートはリンク容量と一致しなければならない。

【００８０】このように公式化していくと、Ｅ（ｎ）の算出は制御問題となることがわかる。この問題はＫｏｌａｒｏｖ及びＲａｍａｍｕｒｔｈｙによる前述のＡＢＲ論文に開示されているように、ＡＢＲサービスについて明示的なレート（ＥＲ）を算出する問題に、ある程度、類似している。しかしながら、スイッチ内でＥＲを実現することは単純化できる。ＤＲＣスケジューラには、単一のコントローラで十分である。レート制御はスイッチ内で局部的に生じるため、フィードバック遅延（図６参照）はサンプリング区間Ｔに比して小さい。この遅延は、フィードバック遅延を考慮する必要があるＡＢＲにおけるフロー制御とは異なり、無視できる。

【００８１】上記した点を考慮すると、Ｋｏｌａｒｏｖ及びＲａｍａｍｕｒｔｈｙの論文に開示された制御はノンリアルタイムＡＢＲサービスに対してのみ適用できるのに対し、本発明に係わる制御は、リアルタイム及びノンリアルタイムのサービスに対して適用できることが特徴である。

【００８２】コントローラの設計を単純化するために、滞留時を無限に有する単一ソースが第１段に設けられているものと仮定する（すなわち、この滞留時により、利用可能なリンク容量が常に満たされているものとする）。Ｅ（ｎ）は時間ｎにおいてコントローラにより算出された超過レートを示すものとする。この場合、Ｒ（ｎ）は、単一ソースに対する第１段から第２段へのフローレートである。

【００８３】ε（ｎ）＝Ｑ（ｎ）－Ｑ０は、時間ｎにおけるキューの長さと目標のキューの長さＱ０との誤差をあらわすものとする。離散的な時間に関するＰＤ（Proportional Differential）コントローラは、一般に、次の数２式によってあらわすことができる。

【００８４】

【数２】

$$E(n+1)=E(n)-\alpha_0\varepsilon(n)\ \alpha_1\varepsilon(n-1)-\cdots$$
$$-\alpha_u\varepsilon(n-u)-\beta_0E(n)-\beta_1E(n-1)-\cdots$$
$$-\beta_vE(n-v) \qquad (1)$$

ここで、αｉ（ｉ＝１，…，ｕ）と、βｉ（ｉ＝１，…，ｖ）は、実数の係数である。

【００８５】ＤＲＣスケジューラでは、単純な２パラメータフィルタを用いることができ、この場合、（１）式は次式のように簡略化できる。

【００８６】

$$E(n+1)=E(n)-\alpha_0\varepsilon(n)-\alpha_1\varepsilon(n-1) \qquad (2)$$

上記したことから、本発明に係わるコントローラは簡略化した構成を備えると共に、レートシェーピングの速度を向上させることができる。

５．３　オーバーロード（過負荷）制御

閉ループコントローラは、誤差ε（ｎ）＝Ｑ（ｎ）－Ｑ０の絶対値が小さくなるようにレートＥ（ｎ）を調整する。平衡状態において、十分なフローがソースから与えられている場合、第２段バッファへの全フローレートはリンク容量と一致しなければならない。

【００８０】このように公式化していくと、Ｅ（ｎ）の算出は制御問題となることがわかる。この問題はＫｏｌａｒｏｖ及びＲａｍａｍｕｒｔｈｙによる前述のＡＢＲ論文に開示されているように、ＡＢＲサービスについて明示的なレート（ＥＲ）を算出する問題に、ある程度、類似している。しかしながら、スイッチ内でＥＲを実現することは単純化できる。ＤＲＣスケジューラには、単一のコントローラで十分である。レート制御はスイッチ内で局部的に生じるため、フィードバック遅延（図６参照）はサンプリング区間Ｔに比して小さい。この遅延は、フィードバック遅延を考慮する必要があるＡＢＲにおけるフロー制御とは異なり、無視できる。

【００８１】上記した点を考慮すると、Ｋｏｌａｒｏｖ及びＲａｍａｍｕｒｔｈｙの論文に開示された制御はノンリアルタイムＡＢＲサービスに対してのみ適用できるのに対し、本発明に係わる制御は、リアルタイム及びノンリアルタイムのサービスに対して適用できることが特徴である。

【００８２】コントローラの設計を単純化するために、滞留時を無限に有する単一ソースが第１段に設けられているものと仮定する（すなわち、この滞留時により、利用可能なリンク容量が常に満たされているものとする）。Ｅ（ｎ）は時間ｎにおいてコントローラにより算出された超過レートを示すものとする。この場合、Ｒ（ｎ）は、単一ソースに対する第１段から第２段へのフローレートである。

【００８３】ε（ｎ）＝Ｑ（ｎ）－Ｑ０は、時間ｎにおけるキューの長さと目標のキューの長さＱ０との誤差をあらわすものとする。離散的な時間に関するＰＤ（Proportional Differential）コントローラは、一般に、次の数２式によってあらわすことができる。

【００８４】

【数２】

$$E(n+1)=E(n)-\alpha_0\varepsilon(n)\ \alpha_1\varepsilon(n-1)-\cdots$$
$$-\alpha_u\varepsilon(n-u)-\beta_0E(n)-\beta_1E(n-1)-\cdots$$
$$-\beta_vE(n-v) \qquad (1)$$

ここで、αｉ（ｉ＝１，…，ｕ）と、βｉ（ｉ＝１，…，ｖ）は、実数の係数である。

【００８５】ＤＲＣスケジューラでは、単純な２パラメータフィルタを用いることができ、この場合、（１）式は次式のように簡略化できる。

【００８６】

$$E(n+1)=E(n)-\alpha_0\varepsilon(n)-\alpha_1\varepsilon(n-1) \qquad (2)$$

上記したことから、本発明に係わるコントローラは簡略化した構成を備えると共に、レートシェーピングの速度を向上させることができる。

５．３　オーバーロード（過負荷）制御

閉ループコントローラは、誤差ε（ｎ）＝Ｑ（ｎ）－Ｑ０の絶対値が小さくなるようにレートＥ（ｎ）を調整する。しかしながら、全入力トラフィックＲ（ｎ）における変動は閉ループコントローラの動作速度も速いことがある。この場合、目標値Ｑ０に近くなる前に、キューの長さＱ（ｎ）が大きな値に増大することがある。このような状態は、最低保証レートＭｉよりかなり大きいレートで伝送を行なうコネクションにより引き起こされる。

Ｑ（ｎ）の値が大きくなると、最低レートに近いレートで伝送を行なっているコネクションの遅延性能に悪影響を及ぼすおそれがある。閉ループコントローラの応答時間は遅すぎてスケジューラの第２段における過負荷を回避することができないため、好ましい実施例では、別の過負荷制御手法が用いられる。

【００８７】第２段バッファが特定のシェーピングしきい値を越えると、フィードバックシェーピング信号がＤＲＣスケジューラに送られる。このシェーピング信号は、スケジューラに、全てのキューを最低保証レートＭｉでシェーピングさせ、未使用帯域幅の分配を停止させる。この動作により、輻輳を緩和できる高速過負荷制御を実現できる。従来技術のバックプレッシャ信号と異なり、本発明に係わるシェーピング信号は、輻輳が存在する場合に、最低保証レートでキューを伝送させる重要な機能を持っている。ここでも、全てのキューに第２段バッファのキューに対して全てのセル送出を停止させることを指示するストップバックプレッシャ信号が更に使用されてもよい。

【００８８】特に、従来技術において、単純な二元的（停／進）バックプレッシャ制御を行なうことにより、全てのキューに対するスループットを等化できる。すなわち、各キューはＣ／Ｎスループットを達成する。さらに、二元的信号がしばしば発生すると、セル遅延変動（ＣＤＶ）が大きくなり、これによって、リアルタイムコネクションのＱｏＳが悪影響を受けるおそれがある。

【００８９】前述した点を考慮して、好ましい実施例では、２つの信号、すなわち、シェーピング信号及びストップ信号が用いられる。シェーピング信号しきい値は、ストップ信号しきい値より低く設定され、輻輳を緩和させるとともに、最低保証レートでの送信が可能とする。ストップ信号しきい値は非常に高く設定され、シェーピング信号を用いていることから、ストップ信号はリリーフ弁のように、最後の手段として出力されるため、殆ど出力されない。

５．４　大容量スイッチのためのＤＲＣスケジューリング

５．４．１　単一ループフィードバック

図２ｂの入力モジュールにおいて、セルは、トラフィッククラスおよびコアスイッチモジュールの宛先出力ポートに応じて、キューとして格納される（以下、このような構成をクラス／ＯＰキューとして示す）。ここで、（ｉ，ｊ，１）は、クラスｉ及び入力モジュール１内の宛先出力ポートｊに対応するキューを示すものとする（以下では、入力モジュール内の特定のキューに言及するために、入力モジュールに番号を付けずに、単に、（ｉ，ｊ）と表記する場合もある）。各クラス／ＯＰキューは、フローレートを制御可能な仮想ソースとして取り扱われる。

【００９０】コネクション受付制御（ＣＡＣ）アルゴリズムにしたがって、各キューに対して最低保証サービスレートを与えるスイッチを構成できる。この最低レートの値は、クラスのＱｏＳ目標と、受付けたコネクション数と、これらのトラフィック特性により決定される。Ｍｉｊ１は、キュー（ｉ，ｊ，１）の最低保証サービスレートを示すものとする。入力モジュールスケジューラでは、各入力モジュールキューが最低保証レート以上のスループットを達成できることを保証しなければならない。この場合、リアルタイムトラフィックに対しては、比較的高速におけるスループットが保証されなければならない。また、ノンリアルタイムトラフィックについても最低保証スループットを確保しなければならないが、この場合には、より低速におけるスループットが保証されなければならない。

【００９１】最低レートの合計は、各出力ポート及び各入力モジュールにおける回線容量を越えてはならない。すなわち、次の数３式の関係を満足しなければならない。

【００９２】

【数３】

$$\sum_{i,1} M_{ij1} \leq C, \quad for \quad each \quad OP \quad j \quad and$$

$$\sum_{i,1} M_{ij1} \leq C \quad for \quad each \quad IM \quad 1.$$

最低レートにしたがって、キューフローを静的にレートシェーピングすることにより、レートを保証することができる。しかしながら、上述したように、静的スケジューリング規則のもとでは、出力ポートボトルネック部において帯域幅を利用できる場合であっても、割り当てられた最低レートより高いレートでキューを伝送することはできない。静的なスケジューリング規則では、出力ポートのボトルネック部に関連して動作が維持されない（すなわち、出力ポートでは、送信すべきキューが入力モジュールに存在する場合でも、不動作の状態になる時間がある）から、キューフローが統計的に多重化されない。この場合、キューフロー内のコネクション間で統計的多重化が行なわれるだけである。

【００９３】キューフロー間で統計的多重化を達成するために、本発明では、動的レート制御（ＤＲＣ）スケジューリングが使用される。ボトルネック部となる可能性のある出力ポートｊにおいて、超過レートＥｊは、出力ポートにおけるトラフィック利用度及び宛先ＯＰｊに対応する入力モジュールの全てのキューにおけるキュー長さをあらわす情報に基づいて算出される。Ｅｊを算出す

るための方法については、後述する。動的レートＲｉｊｌは、次式にしたがって、キュー（ｉ，ｊ，１）に割り当てられる。

【００９４】Ｒｉｊｌ＝Ｍｉｊｌ＋ｗｉＮｉｊｌ Ｅｊ

ここで、ｗｉは、クラスｉに対して予め割り当てられた重み係数であり、Ｎｉｊ１は、キュー（ｉ，ｊ，１）に関連するアクティブコネクションの数をあらわす。リアルタイムコネクションについて言えば、Ｎｉｊ１は、単純にキュー（ｉ，ｊ，１）に割り当てられたコネクションの数である。ノンリアルタイムコネクションについては、Ｎｉｊ１は、いくつかの実際のコネクションが不動作状態にあるため、キュー（ｉ，ｊ，１）に割り当てられたアクティブコネクション数の推定値として算出される。ｗｉの値は、マルチクラスＣＡＣに用いられる手法に基づいて決定される。

【００９５】上記したことからも明らかな通り、レートＲｉｊ１は、ＣＡＣにより決定される静的部分Ｍｉｊ１に対して、動的部分ｗｉＮｉｊ１Ｅｊを加えたものである。静的部分により、最低保証レートが与えられ、一方、動的部分により、キューフローが、出力ポートのボトルネック部において、公平に（重みｗｉにより決定される）且つボトルネック部にオーバーロードすることなく、未使用帯域幅を有効に利用することが可能になる。

【００９６】ＤＲＣスケジューリングは、また、出力モジュールキュースケジューラに対しても同様に適用される。出力モジュールにおいて、セルは、トラフィッククラス及び出力回線に応じて、キューイングされる。入力モジュールスケジューリングについて前述したものと同じ表記を用い、（ｉ，ｊ，１）が、クラスｉ及び出力モジュールＯＭ１における宛先出力回線ｊに対応するキューを示すものとする。ＯＭ１におけるキュー（ｉ，ｊ，１）に割り当てられた動的レートＲｉｊ１（チルダ付き）は、次の数４式によって決定される。

【００９７】

【数４】

$$\tilde{R}_{ij1} = \tilde{M}_{ij1} + w_i \tilde{N}_{ij1} \tilde{E}_{j1}$$

ここで、Ｍｉｊ１（チルダ付き）は、キュー（ｉ，ｊ，１）についての最低保証レート、Ｎｉｊ１（チルダ付き）は、キュー（ｉ，ｊ，１）に割り当てられたアクティブコネクションの数、Ｅｊ１（チルダ付き）は、ＯＭ１における回線ｊについてのＤＲＣレートである。

５．４．２　デュアルループフィードバック

前述したセクションにおいて、ＤＲＣレートＥｊは、出力ポートＯＰｊにおけるボトルネック部について算出され、当該ＤＲＣレートＥｊがレートＲｉｊ１を算出するのに用いられた。第２のボトルネック部として、レートの不一致により、出力モジュールの出力回線にボトルネックが生じる場合が考えられる。出力回線がオーバーロード状態になると、出力モジュールのキューは長くなり、セル損失が多くなると共に過少利用率の状態になることがある。

【００９８】このような問題は、図２ｃに示すように、出力モジュール内の出力回線に応じて入力モジュールにおけるセルをキューイングすることにより、緩和することができる。

【００９９】より具体的に言えば、（ｉ，ｊ，ｋ，１）の表記は、クラスｉ、宛先出力モジュールｊ、宛先出力回線ｋ（出力モジュールｊ内）、及び入力モジュール１のキューを示すものとする。また、表記（ｉ，ｊ，ｋ）は、宛先ＯＭｊ及び出力回線ｋにおけるクラスｉのキューをあらわすものとする。この場合、入力モジュール内のキューの数は係数Ｌと共に増加する。ここで、Ｌは出力モジュール毎の出力回線の数である。入力モジュールから出力回線のボトルネック部に第２のフィードバックループを追加して、ＤＲＣスケジューリングを行なうことにより、スイッチ性能をさらに向上させることができる。以下、この点を簡単に説明する。

【０１００】第２のフィードバックループにおいて、レートＥｊｋはＯＭｊ内の出力回線ｋにおけるキュー内のセルの数に基づいて算出される。ここで、Ｅｊｋは出力モジュールｊ内の出力回線ｋに対応するボトルネック部において利用可能な帯域幅をあらわし、この帯域幅をレートとして、全ての入力モジュールＩＭに伝達される。これを用いて、各入力モジュールではキュー（ｉ，ｊ，ｋ，１）に対する動的レートＲｉｊｋｌを次式により算出する。

【０１０１】

Ｒｉｊｋｌ＝Ｍｉｊｋｌ＋ｗｉｍｉｎ｛Ｅｊ、Ｅｊｋ｝

ここで、Ｍｉｊｋｌは、キュー（ｉ，ｊ，ｋ，１）についての最低保証レートを示す。このようにして動的レートを算出することにより、出力モジュールｊ及び出力回線ｋ（出力モジュールｊ内）における双方のボトルネック部のレートを動的に調整できる。実際には、入力モジュールＩＭのキューを制御する動的レートは、前述した２つのフィードバックループにおいて算出される。これら２つのフィードバックループのうち、第１のフィードバックループは任意の出力モジュールに接続されており、第２のフィードバックループは出力モジュール内の任意の出力同様に接続されている。この構成では、出力モジュールにおけるキューを短くし、且つ、高い利用率を維持することができる。すなわち、ＤＲＣスケジューリングによる制御を行なう場合、セルをキューイングする動作は大部分入力モジュールで行なわれることになる。

【０１０２】６．スケジューラの設計

６．１　レートベースのスケジューリング

各セル時間内に、入力モジュールＩＭのスケジューラは、セルを宛先ＯＰに送出できる次のキューを決定する。ＤＲＣスケジューリングは、各キューに対して動的

に算出されたレートに基づいて行なわれる。この動的レートを与えられると、入力モジュールＩＭは、処理すべき次のキューを決定する２段階アルゴリズムを実行する。

【０１０３】● 第１段：　仮想レートシェーピング
● 第２段：　サービススケジューリング仮想レートシェーピング及びサービススケジューリングについては、後述する。

【０１０４】表記（ｉ，ｊ）であらわされるキューについて考察する。この場合、当該キューは、クラスｉ（ｉ＝０，…，７）及び宛先ＯＰｊ（ｊ＝０，…，１５）によってあらわされる。当該キュー（ｉ，ｊ）は、次に示すパラメータを有している。

【０１０５】● ＴＳｉｊ：　タイムスタンプ。これは、１２ビットの整数部及び８ビットの小数部を持つ２０ビットレジスタに記憶される。タイムスタンプは、キューがサービスに対してスケジューリングあるいは再スケジューリングされるたびに、更新される。ＴＳｉｊはゼロに初期化される。

【０１０６】● ＡＱｉｊ：　実際のキューサイズ。これは、キュー内（ｉ，ｊ）に含まれるセルの数である。ＡＱｉｊは、セルがキュー（ｉ，ｊ）に到着するたびに増加し、あるセル時間に処理すべきキューｉをスケジューラが選択するたびに減少する。ＡＱｉｊは、１６ビットレジスタに記憶され、ゼロに初期化される。

【０１０７】● ＶＱｉｊ：　仮想キューサイズ。これは、キュー（ｉ，ｊ）に対する仮想キューに含まれるセルの数である。ＶＱｉｊは、８ビットレジスタに記憶され、ゼロに初期化される。好ましい実施例において、ＶＱｉｊは、２５５を越えて増加することはない。

【０１０８】● ＷＦｉｊ：　ラップアラウンドフラグ。これは、現在クロック（時刻）のサイクルを示す２ビットフラグであり、１に初期化される。

【０１０９】● Ｍｉｊ：　最低保証レート。この量はＣＡＣにより与えられ、間隔ＩＭｉｊとして記憶される。

【０１１０】● Ｊｉｊ：　最低保証レートについての間隔。これは、キュー（ｉ，ｊ）についての最少保証レートＭｉｊの逆元であり、１２ビットの整数部及び８ビットの小数部を持つメモリに２０ビットの数として記憶される。

【０１１１】● Ｅｊ：　ＤＲＣレート。ＤＲＣレートはＯＰｊのグローバル仮想キューに基づきＯＰｊにおいて算出される。この値は、ＩＭに記憶されない。

【０１１２】● ｗｉ：　クラスｉについての重み係数。これは、シェーピングレートＲｉｊの算出に用いられる８ビットの整数である。

【０１１３】● Ｒｉｊ：　算出されたシェーピングレート。これは、キュー（ｉ，ｊ）のトラフィックをシェーピングするためのＤＲＣアルゴリズムによって算出されたレートである。この値は、間隔Ｉｉｊとして記憶される。

【０１１４】● Ｉｉｊ：　算出されたレートの間隔。これは、キュー（ｉ，ｊ）についてＤＲＣアルゴリズムにより算出された、レートＲｉｊの逆元である。

【０１１５】● Ｐｉｊ：　スケジューリング優先権。これは、１ビットフラグである。Ｐｉｊ＝１は、キュー（ｉ，ｊ）が、仮想キューにスケジューリングする次のセルを見つけることの優先権を有することを示す。

【０１１６】● ＰＶｉｊ：　仮想キューからのサービス優先権。これは１ビットフラグである。ＰＶｉｊ＝１は、キュー（ｉ，ｊ）が、仮想キューからサービスする（ＯＰに送出する）次のセルを見つけることの優先権を有することをあらわしている。

【０１１７】● Ｓｉｊ：　シェーピング信号。この信号を設定した場合、キュー（ｉ，ｊ）は、ＭＣＲレートでスケジューリングしなければならない。信号Ｓｉｊは、次の条件において、１に設定される。

【０１１８】１．　キュー（ｉ，ｊ）は、ＲＴタイプであり、シェーピングＲＴ信号は、１に設定される。

【０１１９】２．　キュー（ｉ，ｊ）は、ＲＴタイプであり、ローカルＶＱカウントがしきい値を越える。すなわち、Σｉ，ｊＶＱｉｊ≧ＴｈＶＱの関係が成り立つ。

【０１２０】３．　Ｎｉｊ×Ｗｉ×Ｅｊの積は０である。

【０１２１】ここで、ＮＲＴトラフィックは、第３の場合においてのみシェーピングされる。

【０１２２】● Ｎｉｊ：　アクティブＶＣの数。ＲＴキューについて、この数は単純に、ＣＡＣにより登録されたＶＣの数である。これは、ＶＣを全てアクティブと仮定するからである。長期にわたって休止状態となり得るＮＲＴキュー（すなわち、ＵＢＲやＡＢＲ）について、アクティブなＮＲＴキューの数は、サブセクション５．５で説明する計数メカニズムにより推定される。

【０１２３】上述した好ましい実施例において、タイムスタンプはキュー毎に割り当てられるが、セル毎にタイムスタンプを割り当てることによりレートシェーピングを実行することもできる。このような実施例において、現在時刻ＣＴ以下のタイムスタンプを有する全てのセルはサービスの資格があるものとする。

現在時刻及びラップアラウンド

各入力モジュールＩＭは、現在時刻を記憶する１２ビットカウンタＣＴを有している。ここで、各時間の単位は、２．４Ｇｂｐｓにおける１セル時間、すなわち、１７５ｎｓである。１サイクルは、212セル時間、すなわち、ＣＴ＝０から始めてＣＴが循環する[wrap around]までに必要とする時間として定義される。ＣＴが一巡する度毎に、各キュー（ｉ，ｊ）に対するＷＦｉｊフラグは、ＷＦｉｊが３未満の場合、１ずつ増加する。この動作は、一セル時間内で行なうことができる。タイムスタ

ンプＴＳｉｊ及びフラグＷｉｊは、合せて、現在時刻に関してキュー（ｉ，ｊ）に対して保持される時間の値を示す。ＷＦのとり得る４つの値の意味が、表２に示されている。

【０１２４】

【表２】

：　ＷＦｉｊの値

ＷＦｉｊ	意味：　ＴＳｉｊは…
０	ＣＴの１サイクル前
１	ＣＴと同一サイクル内
２	ＣＴの１サイクル後
３	ＣＴの少なくとも２サイクル後

ＴＳｉｊの値は、ＷＦｉｊとともに、ラップアラウンドが生じた時でさえも、キュータイムスタンプと現在時刻の相対値を決定することを可能とする。

６．２　仮想レートシェーピング

仮想レートシェーピングは、各キュー（ｉ，ｊ）に割り当てられたタイムスタンプＴＳｉｊに基づいて行なわれる。タイムスタンプＴＳｉｊは、キュー（ｉ，ｊ）のフローレートがＲｉｊかＭｉｊに制限されるように、更新される。動的レートＲｉｊは、輻輳が生じていない状態では、タイムスタンプ計算に用いられる。輻輳が生じている状態では、最低レートＭｉｊが用いられる。最低レートをキュー（ｉ，ｊ）に対するタイムスタンプ計算に用いられた場合には、優先ビットＰｉｊが１に設定される。

【０１２５】現在時刻以下のタイムスタンプを有するキューは、有効と判断される。現在時刻は、フリーランニングクロックにより指示されており、クロック間の間隔は１セル時間に等しい。各セル時間の間、入力モジュールＩＭスケジューラは、１に設定された優先ビットＰｉｊを有するキュー（ｉ，ｊ）に与えられる優先権を備えた次の有効キューを選択する。有効キュー、すなわち、キュー（ｉ，ｊ）が選択されると、仮想キューカウンタＶＱｉｊは１だけ増加する。仮想キューカウンタは、キューとして有効であるが、まだ送出されていないセルの数を計数する。ＡＱｉｊはキュー（ｉ，ｊ）内のセルの数を示す。仮想キュー内にスケジュールされていないアクティブなキューにまだセルがある場合、すなわち、ＡＱｉｊ＞ＶＱｉｊであれば、タイムスタンプＴＳｉｊは、実際のキュー内の次のセルをいつスケジューリングするかを指示するように更新される。タイムスタンプはまた、キュー（ｉ，ｊ）に、セルが到来すると更新される。

【０１２６】スケジューリングアルゴリズムにしたがって、次の２つの条件により、優先ビットＰｉｊが設定される。

【０１２７】１．　タイムスタンプＴＳｉｊの現在時刻からの遅れが１／Ｍｉｊよりも多い。

【０１２８】２．　ＯＰｊにおける輻輳を制御するために、最低保証レートＭｉｊでキューフローをシェーピングする必要がある。第１のケースは、ローカルＩＭスケジューラがトラフィック及び算出されたシェーピングレートを維持することができないために、発生する。レート計算が適切なレートに収束するのに時間がかかるため、ＩＭスケジューラにおけるシェーピングレートの瞬時的合計が回線レートＣを超過してしまう場合がある。これは、次の数５式であらわされる。

【０１２９】

【数５】

$$\sum_{ij} R_{ij} > C$$

この状態が発生すると、複数のキューが同じセル時間内に有効となる。

【０１３０】第２のケースにおいて、キューは、シェーピングモードでスケジューリングされる。すなわち、キューサービスレートがＲｉｊからＭｉｊに変更される。シェーピングモードは、以下の条件のもとで、呼出される。

【０１３１】１．　キューはリアルタイムであり、ＯＰからのシェーピング信号あるいはストップ信号が有効である。

【０１３２】２．　キューがリアルタイムで、宛先ＯＰについての仮想キューカウンタＶＱのカウント値の部分合計値がしきい値を越える。

【０１３３】３．　算出された動的レートＲは、最低レートＭに等しい。

【０１３４】最初の２つの条件は、出力ポートにおけるリアルタイムトラフィックについて輻輳の発生を示すものである。この状態ではシェーピングモードに迅速に切り換えることにより、このような輻輳を緩和するとともに、最低レート保証を維持できる。シェーピングモードで動作するキューに優先権を与えることにより、最低レートを保証できる。有効なキュー（ｉ，ｊ）が第１段スケジューラにより選択されると、その仮想キューカウンタＶＱｉｊが１だけ増加する。優先ビットＰｉｊが１に設定されると、第２段の優先ビットＰＶｉｊが１に設定される。

【０１３５】セルが空のキュー（ｉ，ｊ）に加わると、キューは有効になり、次にキューが処理される時間を指示するためにタイムスタンプＴＳｉｊが再計算される。このプロセスをスケジューリングと呼ぶ。第１段スケジューラがキュー（ｉ，ｊ）を選択した後、仮想キューカウンタＶＱｉｊは１だけ増加する。アクティブなキュー内に、仮想キュー内にスケジューリングされていないセルがまだある場合、すなわち、ＡＱｉｊ＞ＶＱｉｊの場合には、タイムスタンプＴＳｉｊは、実際のキュー内の次のセルがスケジューリングされる時間を指示するように更新される。このプロセスを再スケジューリングと呼ぶ。スケジューリングスケジューリングについてのアル

ゴリズムをフローチャート形式で図７に示す。ステップＳ７００において、キュー（ｉ，ｊ）のセルが到着すると、カウンタＡＱｉｊが１だけ増加する（＋＋の表記は、１だけ増加することを示すために用いられる）。ステップＳ７１０において、ＡＱｉｊ－ＶＱｉｊ＝１が成立するかどうかチェックされる。成立する場合には、キューのスケジューリングが可能となり、プロセスはステップＳ７２０に移行する。成立しない場合には、処理はステップＳ７２５で終了する。

【０１３６】可変ＣＣＴは、タイムスタンプＴＳｉｊに関連する現在時刻の値を含む１４ビットの整数変数である。ＷＦｉｊは１に初期化される。表２を参照すると、ＡＱｉｊ－ＶＱｉｊ＝１が成立する場合には、必然的に、ＷＦｉｊ≧１となることがわかる。

【０１３７】ステップＳ７２０において、ＣＣＴは、次式により計算される。

【０１３８】ＣＣＴ＝ＣＴ＋（ＷＦｉｊ－１）＜＜１２

ここで、＜＜は、２進左シフト（桁送り）演算（すなわち、212の乗算）を示している。

【０１３９】スケジューリングアルゴリズムにおける次のステップは、ＣＣＴと１３ビットの整数ＴＳｉｊ＋ＩＭｉｊ″を比較することである。ステップ730において、

ＣＣＴ＜ＴＳｉｊ＋Ｊｉｊ″　　（３）

なる条件が成立しない、即ち、２偽（フォールス）であるならば、キュー（ｉ，ｊ）の処理が遅いと考えられる。すなわち、最低保証レート（Ｊｉｊ）において少なくとも１区間だけ現在時刻より遅れている。このため、処理はステップＳ７４０に進み、キュートラフィックがその最低保証レートに従っているため、優先ビットＰｉｊは１に設定される。次に、ステップＳ７５０において、キュー（ｉ，ｊ）は、現在時刻ＣＴにスケジューリングされる。すなわち、ＴＳｉｊ＝ＣＴとなる。

【０１４０】しかしながら、ステップＳ７３０において、式（３）が真であるならば、処理はステップＳ７６０に進んで、Ｓｉｊの値をチェックする。Ｓｉｊの値に応じて、最低保証レートＭｉｊあるいは算出されたレートＲｉｊでキューは、スケジューリングされる。ステップＳ７６０において、Ｓｉｊ＝０であれば、すなわち、ＯＰがオーバーロードされていなければ、キュー（ｉ，ｊ）には、ステップＳ７７０において、優先権Ｐｉｊ＝０を付すると共に、算出されたレートＲｉｊが割り当てられる。この場合、ステップＳ７８０において、

ＣＣＴ＜ＴＳｉｊ＋Ｉｉｊ″　　（４）

の条件が偽であるならば、処理はステップＳ７５０に進み、キューは現在時刻ＣＴにスケジューリングされる。ステップＳ７８０の条件が真であれば、タイムスタンプは、ステップＳ７９０において、次式のように更新される。

【０１４１】ＴＳｉｊ＝ＴＳｉｊ＋Ｉｉｊ

ステップＳ７３０において、Ｓｉｊ＝１であると、すなわち、ＯＰがオーバーロードになると、キュー（ｉ，ｊ）には、最低保証レートシェーピングされ優先権Ｐｉｊ＝１が与えられる（ステップＳ７７５）。この場合、ステップＳ７９５において、タイムスタンプは次式にしたがって更新される。

【０１４２】ＴＳｉｊ＝ＴＳｉｊ＋Ｊｉｊ

また、ラップアラウンドフラグＷＦｉｊは、ステップＳ７５５において、適切に調整される必要がある。

再スケジューリング

再スケジューリングアルゴリズムを図８に示す。このアルゴリズムは、現在時刻ＣＴに対して適合時刻を過ぎてしまったキューを処理し、必要に応じて再スケジューリングすることを目的としている。ここで、キュー（ｉ，ｊ）を処理するということは、その仮想キューカウンタＶＱｉｊを１だけ増加することを意味する。アルゴリズムは、Ｐｉｊ＝１の優先権を有するキューに（ｉ，ｊ）のラウンド・ロビンサーチを行なう。この場合、サーチは、クラスｉ（ｉ＝０，…，７）と、さらに宛先ＯＰｊ（ｊ＝０，…，１５）に対して反復的に行なわれる。図８には示されていないが、このサーチは以下のようにして、実行される。まず、第１のサーチ処理では、Ｆｉｊ＝１で且つ、優先ビットがセットされた（即ち、Ｐｉｊ＝１）キュー（ｉ，ｊ）を見つける動作が行なわれる。このサーチでキューが見つからなかった場合、第２のサーチ処理が、Ｆｉｊ＝１をもつキューを見つけるために実行される。再スケジューリングアルゴリズムは、全てのキューが検査されるか、現在セル時間の終了を示すタイムアウトが起こるまで、実行される。

【０１４３】Ｆｉｊ＝１という条件は、下記が成立する場合にのみ真である。

【０１４４】１．　ＡＱｉｊ＞ＶＱｉｊ。これは、実際のキュー内に、仮想キュー内にスケジューリングされていない少なくとも１つのセルがあることを意味する。

【０１４５】２．　ＶＱｉｊ＜ＦＦ（１６進）。カウンタＶＱｉｊは、ＶＱｉｊ＝ＦＦ（１６進）であるとき、増加を停止する８ビットカウンタである。したがって、最大２５６個のセルを仮想キューにスケジューリングできる。最大値に達すると、再スケジューリングのためにキューはバイパスされなければならない。すなわち、仮想キューカウンタは２５５を越えて増加することはできない。仮想キューカウンタの限界を１に設定した場合、これは、実際には、仮想キューを使用不能にすることになる。スケジューラは、レートシェーピングを行なうが、レート計算は、仮想キューの全体のサイズに基づくものであってはならない。

【０１４６】３．　ＷＦｉｊ≧２あるいは（ＷＦｉｊ＝１及びＴＳｉｊ≧ＣＴ）。この条件が真であるならば、キュー（ｉ，ｊ）はその適合時間を通過したこと、すなわち、ＴＳｉｊは、ＣＴに記録されている現在時刻より

早い時点をあらわしている。有効なキュー（ｉ，ｊ）が、ラウンド・ロビンループサーチ中に見出だされると、次に、ＶＱｉｊを１だけ増加させる処理が行なわれる（Ｓ８２０）。仮想キューの優先ビットは、次式にしたがって更新される（Ｓ８２０）。

【０１４７】ＰＶij＝ｍａｘ（Ｐij，ＰＶij）すなわち、ＰＶｉｊは、それがすでに１に設定されていたか、あるいは、Ｐｉｊが設定されている場合に１に設定される。

【０１４８】次に、ＡＱｉｊ＞ＶＱｉｊであれば（Ｓ８３０）、キューを再スケジューリングする必要があり（Ｓ８４０）、そうでなければ、再スケジューリングの必要はない（Ｓ８１５）。再スケジューリングステップにおいて、一時的な変数であるＣＣＴがスケジューリングアルゴリズムと同様に算出される（図７参照）。

【０１４９】ＣＣＴ＝ＣＴ＋（ＷＦij－１）≪１２ＣＣＴ≪ＴＳｉｊ＋Ｊｉｊ″が偽であれば、キュー（ｉ，ｊ）は、現在時刻に遅れていると判定される（Ｓ８５０）。したがって、現在時刻に追いつくために、キューは、最低保証レートＭｉｊで、１に設定された優先ビットＰｉｊを付して、スケジューリングされる（ステップＳ８６５及びＳ８７５）。一方、ＣＣＴ≪ＴＳｉｊ＋Ｊｉｊ″が真であれば、Ｓｉｊの値はステップＳ８６０においてテストされる。Ｓｉｊ＝０のとき、キューは、Ｐｉｊ＝０を付してレートＲｉｊでスケジューリングされる（ステップＳ８７０及びＳ８８０）。そうでない場合、処理は、ステップＳ８６５及びＳ８７５に進み、キューは、優先権Ｐｉｊ＝１が付され、最低保証レートでスケジューリングされる。

【０１５０】ＴＳｉｊがＪｉｊあるいはＩｉｊを加えることにより更新されると、その結果、オーバーフロービットＺｉｊが生じる。Ｚｉｊ＝１の場合、タイムスタンプＴＳｉｊは次のサイクルに進み、ＷＦｉｊは１だけ減少されることになる。そうでなければ、ＷＦｉｊはそのままの状態を維持する。この動作は、ステップＳ８９０で行なわれる。

６．３　サービススケジューリング

各セル時間中には、高々１個のセルが、入力モジュールからコアＴＤＭバス上を宛先出力ポートに送出される。図９に示されるように、セルを送り出すべきキューは、優先ビットＰＶｉｊに基づいて、ラウンド・ロビンサーチにより、決定される（ステップＳ９００）。第２段スケジューラにおいて、キューは、ＶＱｉｊ＞０で、且つ、キューについての宛先ＯＰバッファがストップモードにない場合、サービスを受け得るものと判定される。サービスを受け得るキューが見つかると（ステップＳ９００でｙｅｓ）と、キュー内の最初のセルがＴＤＭバス上を伝送される（ステップＳ９１０）。また、ステップＳ９１０において、ＶＱｉｊ及びＡＱｉｊの両方が１だけ減少し、仮想キュー優先ビットＰＶｉｊがゼロにリセットされる。再び、ＶＱｉｊの値は、サービスを受け得るセルの数を示していることは注意すべきである。

６．４　ハードウェア構成

図１０は、スケジューリング動作を実行するためのハードウェアの構成を示している。主要な構成要素は、以下のとおりである。

【０１５１】１．　タイムスタンプＴＳｉｊ用の記憶装置。これらは、複数個のレジスタ１００として実現することができる。

【０１５２】２．　コンパレータ１１０のアレイ。キュー（ｉ，ｊ）に関連づけられたコンパレータは、ＴＳｉｊとＣＴを比較する。

【０１５３】３．　実際のキューＡＱｉｊ用の記憶装置。これはカウンタ２００のアレイとして実現できる。

【０１５４】４．　仮想キューＶＱｉｊ１３０用の記憶装置。

【０１５５】５．　コンパレータの出力に対して優先権をラウンド・ロビン（ＰＲＲ）式にサーチするブロック１３５（仮想レートシェーパー）。

【０１５６】６．　仮想キューについてＰＲＲサーチを実行するブロック１４５（サービススケジューラー）。

【０１５７】７．　計算エンジン１５０。

【０１５８】８．　コアスイッチからの「ｓｔｏｐ／ｓｈａｐｅ／ｇｏ」信号。

【０１５９】図示された仮想レートシェーパー１３５は、キュー（ｉ，ｊ）に対する優先ビットＰｉｊを用いて、仮想レートシェーピングを行なう。仮想レートシェーピングの際、仮想レートシェーパー１３５は、Ｐｉｊ＝１の優先権をもち、ＴＳｉｊ≦ＣＴが成立するキュー（ｉ，ｊ）を探す。ＡＱｉｊ＞ＶＱｉｊであれば、仮想キューＶＱｉｊは、１だけ増加する。

【０１６０】サービススケジューリングのためのＰＲＲサーチを行なうサービススケジューラ１４５は、ＰＶｉｊ＝１の優先権を持ち、ＶＱｉｊ＞０の仮想キューをラウンド・ロビン式にサーチする。仮想キュー（ｉ，ｊ）は、宛先出力ポートｊに対応するストップ信号（ｓｔｏｐ）がない場合にのみ、有効となる。

【０１６１】計算エンジン１５０は、ＤＲＣスケジューリングに応じて、レートＲｉｊを動的に更新する。レートＲｉｊは、次式にしたがって算定される。

【０１６２】Ｒij＝Ｍij＋ｗｉＮijＥｊ

ここで、式中の値は次のとおりである。

【０１６３】●　ＣＡＣからの情報：

－　最低保証レートＭｉｊ

－　クラス重みｗｉ

●　コアスイッチモジュールからのストップ／ゴー／シェープフィードバック。

【０１６４】●　キュー（ｉ，ｊ）に関連づけられた、アクティブなコネクションの推定数Ｎｉｊ。

【０１６５】●　出力モジュールｊからのＩＲＭセル中

に含まれる超過レートＥｊ。計算エンジンは、さらに、サブセクション６．２で説明したスケジューリング及び再スケジューリングアルゴリズムにしたがって、タイムスタンプＴＳｉｊを更新する。

６．５　アクティブＶＣの推定数

キュー（ｉ，ｊ）に対するアクティブＶＣの数Ｎｉｊは、レートＲｉｊの計算と、出力モジュールで算出されるＥＲ値に用いられる。リアルタイムコネクションについては、アクティブＶＣの数は、ＣＡＣアルゴリズムにより受け付けたＶＣの数であると見なされる。ＵＢＲやＡＢＲなどのノンリアルタイムコネクションについては、ＣＡＣによって受け付けたＶＣの数は、任意の時間におけるアクティブなＶＣの実際の数よりはるかに大きい場合がある。これは、一般に、ノンリアルタイムＶＣではＱｏＳを保証する必要がなく、長期間にわたって空き状態にある場合もあるからである。

【０１６６】そこで、ノンリアルタイムトラフィックについてＶＣ数を推定する方法が必要とされる。好ましい実施例の４０Ｇスイッチでは、単純なＶＣテーブルルックアップ法が用いられる。このテーブルには、各ノンリアルタイムＶＣに対して、１ビットエントリ（ゼロに初期化される）と、キュー識別子（ｉ，ｊ）とが保持されている。時間は、長さＴｓをもつ区間に分割される。ＶＣｋに属するセルがある区間において到着する場合、対応するテーブル上のエントリがゼロであると、このエントリがセット状態となり、カウントＮｉｊは１だけ増加する。あるいは、テーブル上のエントリがすでにセットされている場合には、何の動作も行なわれない。当該区間の終了の際、Ｎｉｊは、その区間のアクティブＶＣの数の推定値をあらわしている。次の区間がスタートする前に、この値Ｎｉｊは全てクリアされる。アクティブＶＣ数のより平滑化された推定値は、指数平均により次の数６式により求められる。

【０１６７】

【数６】

$$\overline{N}_{ij}(n) = \epsilon N_{ij}(n) + (1-\epsilon)\, N_{ij}(n-1)$$

ここで、$\varepsilon \in (0, 1)$。

７．レート計算

７．１　ＤＲＣレート

７．１．１　単一フィードバックループ

単一フィードバックループの一般的構造を図２ｂに示す。それぞれ入力モジュールＩＭ及び出力モジュールＯＭにおいてＤＲＣスケジューリングを行なうためのレート値Ｅｊ（出力モジュールｊに対応）とＥｊ（チルダ付き）（出力回線ｊに対応）は、０．５ｍｓごとに１回計算される（セクション５参照）。

【０１６８】ここでは、ＤＲＣレートＥの算出方法について説明するが、Ｅ（チルダ付き）も同様にして算出される。図１１は、ＤＲＣスケジューリングレートの計算のためのフローチャート図を示す。図１１において、Ｅ（ｎ）は、ｎ番目（０．５ｍｓ）のサンプリング区間において算出された一般的なＤＲＣレート値をあらわす。記号ＶＳ（ｎ）は、ボトルネック部に対応する仮想キューサイズの合計をあらわす。ＤＲＣ値Ｅｊに関連したＶＳ（ｎ）は、全ての入力モジュールを介して、出力ポートｊに送られる全ての仮想キューの合計をあらわしている。同様に、ＮＳ（ｎ）は、全ての入力モジュールを介して、出力モジュールｊに送られるアクティブＶＣの総数をあらわす。

【０１６９】ＤＲＣ値Ｅｊｋに関連したＶＳ（ｎ）は、出力モジュールｊと出力回線ｋに対応する全ての仮想キュー合計をあらわす。この場合、ＮＳ（ｎ）は、出力モジュールｊ内の出力回線ｋに送られるアクティブＶＣの数を示している。

【０１７０】この例では、ＯＰボトルネック部における仮想キューの全体の長さに基づいて、Ｅを算出するために閉ループ比例微分制御器（コントローラ）が用いられる。ＯＰチャネル利用度が値Ｕ０（９５％）を越える（ステップＳ１１１０参照）と、ＯＰボトルネック部に対応する仮想キューの全長を目標値Ｎ０の近くに保つように、制御器によりＥの値を調整する。ＯＰチャネル利用度がＵ０より低い場合、制御器は、利用度がＵ０に近づくようにＥを調整する。

【０１７１】ＣＴ（ｎ）は、ｎ番目の走査期間中にＯＰの出力において観察されるセルの数のカウント値を示す。Ｃが１つの走査期間中のセル時間の数であれば、ｎ番目の期間における利用度はＵ（ｎ）＝ＣＴ（ｎ）／Ｃで算出される（ステップＳ１１００）。Ｖ（ｎ）は、ｎ番目の期間中のＯＰに対応する仮想キュー全長を全てのＩＭついて合計した値を示す。Ｕ（ｎ）＞Ｕ０であるならば、誤差は次式により算出される。

【０１７２】Ｄ（ｎ）＝Ｖ（ｎ）－Ｎ０

ここで、Ｎ０は、目標仮想キューの全長である。さもなければ、誤り信号が目標の利用度Ｃ０＝Ｕ０Ｃに基づいて算出され、誤り信号は次式により算出される。

【０１７３】Ｄ（ｎ）＝ＣＴ（ｎ）－Ｃ０

各走査期間中に、ボトルネックレートが下記の比例微分（ＰＤ）制御に関する式を用いて算出される。これは、誤差をゼロにすることを目的としている（ステップＳ１１４０）。

【０１７４】Ｅ（ｎ＋１）＝Ｅ（ｎ）－α０Ｄ（ｎ）－α１Ｄ（ｎ－１）

上式において、係数α０及びα１は、システム安定性と高速の応答時間を保証するように設計された定数である。本発明者らにより実施されたシミュレーション実験

において、定数は、α０＝１．２５及びα１＝－０．７５に設定された。レートがゼロより大きくなければならないという条件は、次式の演算により確保される。

【０１７５】

Ｅ（ｎ＋１）＝ｍａｘ｛Ｅ（ｎ＋１），０｝

レート値は、また、ボトルネック回線レートによっても限定されなければならない。これはすなわち、次式であらわされる。

【０１７６】

Ｅ（ｎ＋１）＝ｍｉｎ｛Ｅ（ｎ＋１），０｝

レートは、［ｃｅｌｌｓ／０．５ｍｓ］の単位で算出される。誤り信号Ｄ（ｎ）、Ｄ（ｎ－１）及び値Ｅ（ｎ）は、次のレート計算のためにＤ（ｎ－１）として格納される（ステップＳ１１５０）。

７．１．２　デュアルループフィードバック

デュアルループを用いてフィードバックを行なう場合、出力回線にしたがって、入力モジュール内にセルをキューイングすることが必要である（図２ｃ）。また、カウンタは各出力回線に対してキューイングされたセル数を保持しなければならない。ここで、ＡＱｊｋは、出力モジュールｊの出力回線ｋに対してキューイングされたセルの数を示すものとする。

【０１７７】この場合、ＤＲＣレートＥｊｋ（出力モジュールｊ内の出力回線ｋに対応する）は、０．５ｍｓ毎に１回算出される（前述したセクション５．４．２参照）。Ｅｊｋの計算は、単一ループの場合に説明したＥｊの計算と同様である。しかしながら、この場合、実際のキューサイズＡＱｊｋは、図１１においてＶＳで示した仮想キューカウント値のかわりに用いられる。キューサイズＡＱｊｋは、次に説明するように、出力モジュールｊ内の出力回線ｋに対するＡＢＲ明示レートの計算にも用いられる。

７．２　ＡＢＲ明示レート７．２．１　出力モジュールボトルネック

ＡＢＲサービスクラスについて、明示レート（ＥＲ）値は実際のＡＢＲクラスキューのサイズに基づいて算出される。ここで説明するＡＢＲレート計算の方法は、上記に引用したＫｏｌａｒｏｖとＲａｍａｍｕｒｔｈｙによるＡＢＲサービスの論文において展開された方法にある程度類似している。但し、より高速の２．４Ｇｂｐｓ回線速度を扱うための修正とスイッチ方式の実現のための修正が加えられている。さらに、ＡＢＲレート計算は、０．５ｍｓ毎に１回実行される。各宛先ＯＰについて、ＥＲ値ＥＲｊ（ｊ＝１，…，１６）が算出される。

【０１７８】図１２には、明示レートＥＲの計算に関するフローチャートが示されている。このフローチャートは、出力モジュールＯＭボトルネック及び出力回線ボトルネックの両方に適用される。Ｃａｂｒ（ｎ）は、ｎ番目の０．５ｍｓ期間中に到着するＡＢＲセルの数を示す。ステップＳ１２００において、ｎ番目の期間中のＡＢＲに対する利用度は次式のとおり算出される。

【０１７９】Ｕａｂｒ（ｎ）＝Ｃａｂｒ（ｎ）／Ｃ

ここで、Ｃは、ボトルネックレートにおける０．５ｍｓ期間内のセル時間の総数である。

【０１８０】ＡＳ（ｎ）は、ｎ番目の期間についてのボトルネック部（出力モジュールあるいは出力回線）に対応する実際のＡＢＲキューのサイズを示す。すなわち、ＡＳ（ｎ）は、あるボトルネック部に向けられた全てのＡＢＲキューについての実際のキューサイズの合計である。期間ｎ－１における値ＡＳ（ｎ－１）がメモリに記憶されている。この状態で、ＡＳ（ｎ）とＡＳ（ｎ－１）の差がしきい値（ステップＳ１２１０において１５０個のセルとして例示されている）を越えたものとする。このことは、ＡＢＲキューが急速に大きくなり、高速制御を用いなければならないことを示している。したがって、ステップＳ１２１５において、ＩＲＲフィルタがアクセスされる。ＩＲＲフィルタは、また、ＡＳ（ｎ）がしきい値Ｔｈｉｇｈを越えたとき（ステップＳ１２２０）、あるいは、フラグＦ＝１である（ステップＳ１２３０）ときにも、アクセスされる。

【０１８１】ステップＳ１２４０において、ＡＢＲトラフィックの利用度が目標より少ないと判断すると、処理は、ステップＳ１２５０に進む。他方、多い場合には、ステップＳ１２４５に進み、低利得フィルタが用いられる。ステップＳ１２５０において、実際のＡＢＲセルの合計が、低い方のしきい値Ｔｌｏｗより少ないと判定されると、処理はステップＳ１２５５に進み、ここで高利得フィルタが用いられる。しきい値Ｔｌｏｗ以上であれば、処理はステップＳ１２４５に戻り、低利得フィルタが用いられる。

【０１８２】図１３は、ＩＲＲフィルタの動作を示す。ＩＲＲフィルタには、ＤＲＣローカルレートＥの何分の一かに等しいＥＲレートが単純に設定される。この場合、次式の関係が成立する（ステップＳ１３１０）。

【０１８３】ＥＲ（ｎ＋１）＝Ｅ（ｎ）／２

ＩＲＲフィルタは、ＡＳ（ｎ）の値がしきい値Ｔｌｏｗより大きいか小さいかに応じてフラグＦを設定あるいはリセットする（ステップＳ１３２０）。

【０１８４】図１２から、Ｆ＝１の場合に、ＩＲＲフィルタはアクセスされることがわかる。これにより、ＡＢＲトラフィックに対しては厳しい制限が加えられることになる。図１３において、誤り信号Ｄ（ｎ－１）は、ＩＲＲフィルタにおいて使用されなくても更新され、記憶される。

【０１８５】図１４は高利得フィルタの動作を示す。主要な制御式は次の数７式のとおりである。

【０１８６】

【数7】ＥＲ（ｎ＋１）＝ＥＲ（ｎ）－α０Ｄ（ｎ）／ＮＳａｂｒ（ｎ）－α１Ｄ（ｎ－１）／ＮＳａｂｒ（ｎ－１）

41

ここで、ＮＳａｂｒ（ｎ）は、あるボトルネックに対応するアクティブＡＢＲであり、ＶＣ全ての合計の推定値であり、ＡＢＲクラス重みＷａｂｒにより重み付けされた値である。フィルタ係数の値は、ローカルＤＲＣフィルタにおけるものと同じである。すなわち、α０＝１．２、α１＝－．７５である。高利得フィルタについて、フィルタ係数は、ＮＳａｂｒによって決定される。

【０１８７】フィルタにおける動作は以下のように行われる。まず、実際のキュー長さと目標長さの差が決定される。ステップ１４１０において、ステップ１４００で計算された差から、高利得フィルタが使用される。ステップ１４２０において、Ｄ（ｎ）は、次回に備えてＤ

42

（ｎ－１）に置き換えられる。ステップ１４３０において、ｍａｘ｛ＥＲ（ｎ＋１），０｝の動作を行なうことによって、ＥＲ（ｎ＋１）が負ではないことが保証される。更に、ｍｉｎ｛ＥＲ（ｎ＋１），Ｅ（ｎ）｝の動作を行なうことによって、ＥＲ（ｎ＋１）がローカルＤＲＣレートＥ（ｎ）以下であることを保証する。ステップ１４４０において、全てのＥＲ値は、時間的にシフトされる。

【０１８８】図１５に示された低利得フィルタにおける制御式は次の数８式のとおりである。

【０１８９】

【数８】

$$ER(n+1)=ER(n)-\alpha 0D(n)-\alpha 1D(n-1)-\beta 0ER(n)-\beta 1ER(n-1)-\cdots-\beta 10ER(n-10) \quad (6)$$

低利得フィルタの係数はＮＳａｂｒによって決定されない。低利得フィルタのための係数値は、表３に示されている。

【０１９０】

【表３】

低利得ＡＢＲフィルタについての係数値

係数	値
α０	０．０６２７
α１	－０．０５４５
β０	０．８８６４
β１	０．０９５５
β２	０．０５４５
β３	０．０１３６
β４	－０．０２７３
β５	－０．０６８２
β６	－０．１０９１
β７	－０．１５００
β８	－０．１９０９
β９	－０．２３１８
β１０	－０．２７２７

低利得フィルタにおける処理動作は、利得式が異なることを除き、高利得フィルタの処理動作と同じである。したがって、図１５に示された低利得フィルタの処理動作の説明はここでは省略する。

７．３　制御情報の伝送

全てのＤＲＣレートとＡＢＲ　ＥＲレートの計算はそれぞれの出力モジュールＯＭにおいて実行される。各走査期間において、各入力モジュールＩＭは、全ての出力モジュールＯＭにキューの長さをあらわす情報を伝送する。この情報は、内部リソース管理（ＩＲＭ）セルと呼ばれる特別の制御セルによって伝送される。これらのセルは入力モジュールＩＭにより生成され、制御シグナリングオーバーヘッドを構成する。

【０１９１】キュー長さ情報に基づいて、各ＯＭｊは、局部的な制御を行ない、この制御ではＤＲＣレートＥｊを算出する一方、ＡＢＲソース制御の際に、明示レート（ＥＲ）ＥＲｊを算出する。ＡＢＲ　ＥＲ値は、ソースの方向に向かうリソース管理（ＲＭ）セルにより、遠隔ＡＢＲソースに伝送される。同様に、出力モジュールＯＭにより生成されたＩＲＭセルは、入力モジュールＩＭにＤＲＣレート情報を伝送するために用いられる。

【０１９２】８．バッファ管理

スケジューラと共に動作する各入力モジュールＩＭ及び出力モジュールＯＭは、バッファ割当ての機能を持つキューマネージャーを備えている。好ましい実施例による大容量スイッチ構造において、出力モジュールＯＭバッファは、出力回線ボトルネック部における競合から生ずる輻輳を扱い、一方、入力モジュールＩＭバッファは、ＯＰボトルネック部における競合から生ずる輻輳を扱う。入力モジュールＩＭ及び出力モジュールＯＭ内のキューマネージャーは、独立したものであるが、同様な構造を備えている。入力モジュールＩＭ及び出力モジュールＯＭ内のセルバッファは、全てのキュー間で共有されるが、最大キューサイズには制限がある。

【０１９３】各キューには、トラフィッククラス及びＱｏＳ要求に基づいて予め割当てられたセル廃棄しきい値が設定されている。廃棄しきい値は、サイズ増加順に下記のとおり挙げられる。

【０１９４】●　ＣＬＰ＝１であるセルを廃棄。

【０１９５】●　初期パケット廃棄（ＥＰＤ）。新たなパケットに属するセルを廃棄。

【０１９６】●　部分パケット廃棄（ＰＰＤ）。全てのセルを廃棄。キューマネージャーは、最低保証レートにシェーピングされている、あらゆるキューフロー中におけるＣＬＰ＝１であるセルを廃棄する。このようにして、ＣＬＰ＝０であるトラフィックには最低保証レートが割り当てられる。

【０１９７】９．性能評価

ＤＲＣスケジューリングの主要な目標は、スイッチにおけるボトルネックレートを整合させて、輻輳回避と、高効率の維持という２つの目的を満足させることである。さらに競合するクラス間で公平に未使用帯域幅を配分することをも目標の一つとしている。このセクションで

は、スイッチ設計におけるＤＲＣスケジューリングの主要な性能特性を強調するために、いくつかの代表的なシミュレーション結果を提示する。

９．１　レート制御の収束

図１６を参照して、コアスイッチモジュール上の同一の出力ポート１に送出される２つのフローが与えられた場合におけるスイッチの動作を説明する。

【０１９８】１．　一定の入力レート０．５８及び最低保証レートＭ１＝０．６を有する、ＩＭ１上のＣＢＲフロー。

【０１９９】２．　一定の入力レート０．９及び最低保証レートＭ２＝０．３を有する、ＩＭ２上のＵＢＲフロー。この場合、ＵＢＲフローに対しては、その最低保証レートが守られないことがある。これは、ＵＢＲソースがネットワーク端において監視されないことから、発生する。対照的に、ＣＢＲソースは、その最低保証レートより低いレートで実際に伝送を行なっている。

【０２００】フローｉのＤＲＣレートは、Ｒｉ＝Ｍｉ＋Ｅで算出される。ここで、Ｅは、閉ループ制御により算出される利用可能な未使用帯域幅である。時刻０において、帯域は空いているから、初期的に、Ｅ＝１である。すなわち、２つのフローが時刻０において同時に生成されると、各フローは、まず回線レート、すなわち、Ｒｉ（０＋）＝１（ｉ＝１，２）で伝送される。

【０２０１】時刻ｔ＝０＋において、ＯＰ１への総フローレートは１．４８である。したがって、ＯＰ１においてバッファリングが行なわれる一方、全仮想キューが入力モジュールにおいて形成される。ＤＲＣメカニズムは、ＤＲＣレートＥを低下させることにより動作を開始する。

【０２０２】図１７は、フローレートＲｉ（ｔ）のグラフを示している。図から、レートが比較的迅速に（約６ｍｓで）定常状態値に収束することがわかる。ＣＢＲフローＲ１（ｔ）は０．５８の帯域幅を用い、ＵＢＲフローＲ２（ｔ）はレート０．９で入力モジュールＩＭ２に入力される。しかし、ＵＢＲフローには、０．３のスループットしか保証されていない。したがって、ＤＲＣレートＥについての正しい値は、０．１２である。このため、レートは次式のように収束する。

【０２０３】Ｒ１（ｔ）→０．７２及びＲ２（Ｔ）→０．４２ＣＢＲフローＲ１（ｔ）はレート０．７２で出力ポートＯＰ１に伝送できるが、入力モジュールＩＭ１にはレート０．５８で入力される。一方、ＵＢＲフローＲ２（ｔ）は、入力モジュールＩＭ２においてレート０．４２にシェーピングされる。ＵＢＲセルは、セルバッファ容量を超過した後、ＩＭにおいて廃棄される。

９．２　リアルタイム対ノンリアルタイム遅延性能

遅延性能を調べるために、フロー２をランダムなオン・オフ期間を有するＵＢＲフローに置き換えることにより、上記の例を修正する。オン・オフ期間は、それぞれ、平均８及び１２［セル時間］を用いて指数的に分散される。オン期間中、フロー２では、一定レート０．９３で入力モジュールＩＭ２にセルが与えられるものとすると、フロー２の平均レートは、０．３７２である。

【０２０４】また、Ｍ２＝０．３８を設定する。シミュレーションから得られた遅延量を表４に示す。平均遅延量は、対応する９８％信頼区間とともに、セル時間の単位で与えられる。ＣＢＲフローがごくわずかな遅延と遅延ジッタを受けることがわかる。

【０２０５】

【表４】

ＣＢＲフロー１及びＵＢＲフロー２についての遅延性能

遅延量［セル時間］	ＣＢＲフロー１	ＵＢＲフロー２
平均遅延	０．９０±０．０６６	２．５４±０．７１
標準偏差遅延	０．０３	９．０９
平均interdeparture	１．５１±０．１１	１．６１±０．１１
標準interdeparture	０．６７	４．２８

表４において、interdepartureは、セルが出力回線から出力されるときのセル間隔であり、例えば、セルＡＡとＢの出力時刻がそれぞれｔ１及びｔ２であるものとすると、interdeparture fは、ｔ２－ｔ１となる。

【０２０６】ここで、フロー１をＣＢＲクラスからＵＢＲクラスに変更する場合を考慮する。ＵＢＲフローとして、フロー１は、フロー２とともに、ＯＰ１のノンリアルタイムバッファにおいてバッファリングされる。この例のシミュレーション結果を表５に示す。表５における両方のフローの平均遅延量は、表４における対応する遅延量に対して増加していることがわかる。特に、ＣＢＲフローとしてのフロー１の遅延量は全て、ＵＢＲフローとしての対応する遅延量よりも著しく良好であることがわかる。この例は、スイッチ構造が、ノンリアルタイムトラフィックに比べてリアルタイムトラフィックに対して提供するよりきびしいＱｏＳ制御を行なうことを証明している。

【０２０７】

【表５】

45　　46

ＵＢＲフロー１及びＵＢＲフロー２についての遅延性能

遅延量［セル時間］	ＵＢＲフロー１	ＵＢＲフロー２
平均遅延	１．７９±０．４２	１．５９±０．２２
標準偏差遅延	５．０	３．９２
平均interdeparture	１．５　±０．１１	１．６２±０．１２
標準interdeparture	１．０３	４．２６

９．３　ＤＲＣ対静的優先権スケジューリング

図１８は、３つのオン・オフ型フローを与えられたスイッチを示している。３つのフローの仕様を表６に示す。各フローは、トラフィッククラス、平均オン・オフ期間、オン期間時のレート、ソース入力モジュール、宛先出力モジュール、最低保証レート（ＤＲＣに対する）に関連づけられている。

【０２０８】

【表６】

３つのフローの仕様

フローNo.	クラス	平均オン	平均オフ	オンレート	ＩＭ	ＯＰ	Ｍ１
１	Ｒｔ－ＶＢＲ	１２	８	０．９	１	１	０．６５
２	Ｎｒｔ－ＶＢＲ	８	１２	０．９３	１	２	０．３
３	Ｒｔ－ＶＢＲ	７	１３	０．９３	２	１	０．３３

フロー１及び３は、リアルタイムＶＢＲフローであり、フロー２は、ノンリアルタイムＶＢＲフローである。さらに、フロー１及び２は、ＩＭ１におけるサービスについて競合する。この例では、ＤＲＣスケジューリングとＩＭ１における静的優先権スケジューリングを比較する。静的優先権は、リアルタイムＶＢＲフロー１に、ノンリアルタイムＶＢＲフロー２より高い優先権を与える。明らかに、フロー１は、この方式のもとで最高の遅延性能を達成する。しかしながら、このことは、フロー２には悪影響を与える。ＤＲＣスケジューリングでは、両方のフローにレート保証を与えることにより調整が行なわれる。

【０２０９】ＤＲＣ及び静的優先権スケジューリングについての遅延結果をそれぞれ表７及び表８に示す。静的優先権のもとでは、フロー１は小さい遅延しか受けない。しかしながら、フロー２の遅延は、比較的大きい。ＤＲＣのもとでは、フロー１の遅延性能は、少しだけ調整されるが、フロー２の遅延性能は著しく向上する。

【０２１０】

【表７】

ＤＲＣスケジューリングのもとでの遅延結果

遅延量［セル時間］	フロー１	フロー２	フロー３
平均遅延	28.0±3.68	25.6±3.62	34.9±5.53
標準偏差遅延	31.24	31.82	63.40
平均interdeparture	1.57±0.11	2.24±0.14	2.11±0.16
標準interdeparture	1.02	2.00	3.14

【０２１１】

【表８】

静的優先権スケジューリングのもとでの遅延結果

遅延量［セル時間］	フロー１	フロー２	フロー３
平均遅延	10.5±1.81	121.5±17.4	3.25±0.33
標準偏差遅延	26.48	104.3	7.72
平均interdeparture	1.56±0.11	1.92±0.14	1.57±0.11
標準interdeparture	2.09	3.68	3.94

【０２１２】

【発明の効果】上記説明からわかるように、本発明によるスイッチは、異なるＱｏＳを有するセルストリームを効率的に処理するものである。さらに、本発明によるスイッチは、効率的な優先権方式を用いて、ユニキャスト及びマルチキャスト伝送を効率的に多重化する。スイッチの入力側と、コアと、出力側にバッファを用いることにより、ボトルネックを悪化させることなく、動作維持を可能とする。さらに、シェーピングフィードバック信号が、一時的な輻輳を緩和させるとともに、最低保証レートを確約するために、動作維持を一時的に停止するために用いられる。

【０２１３】また、本発明では、入力バッファ、出力バッファ、及び、コアバッファの３重バッファ形式を採用すると共に、フィードバック方式を採用することにより、高速の大容量バッファを必要とせず、且つ、入力ポート間における制御を軽減できる。

【図面の簡単な説明】

【図１】本発明の好ましい実施例によるコアスイッチモジュールの構造を示す図である。

47

【図２】（ａ）は、本発明の好ましい実施例によるＶＣ毎のキューを構成する入力及び出力モジュールの構造を示す図である。（ｂ）は、本発明の好ましい実施例による出力ポートクラス毎のキューを構成する入力及び出力モジュールの構造を示す図である。（ｃ）は、本発明の好ましい実施例による出力回線クラス毎のキューを構成する入力及び出力モジュールの構造を示す図である。

【図３】本発明の好ましい実施例による入力及び出力モジュールの構成をより詳細に示す図である。

【図４】最小レートシェーピングをともなうスケジューラを示す図である。

【図５】本発明によるＤＲＣスケジューリングを実現するスケジューラを示す図である。

【図６】閉ループレート制御を示す図である。

【図７】タイムスタンプに応じてセルをスケジューリングするためのアルゴリズムのフローチャートである。

【図８】タイムスタンプに応じてセルを再スケジューリングするためのアルゴリズムのフローチャートである。

【図９】仮想キューからセルを処理するフローチャートである。

【図１０】本発明の好ましい実施例によるスケジューラのブロック図である。

【図１１】ＤＲＣスケジューリングについてレート計算のためのアルゴリズムのフローチャートである。

【図１２】ＡＢＲについての明示的レート計算のフローチャートである。

48

【図１３】ＡＢＲについてのＩＲＲフィルタのフローチャートである。

【図１４】ＡＢＲについての高利得フィルタのフローチャートである。

【図１５】ＡＢＲについての低利得フィルタのフローチャートである。

【図１６】コアスイッチの出力ポートに与えられる２つのセルストリームフローを示す図である。

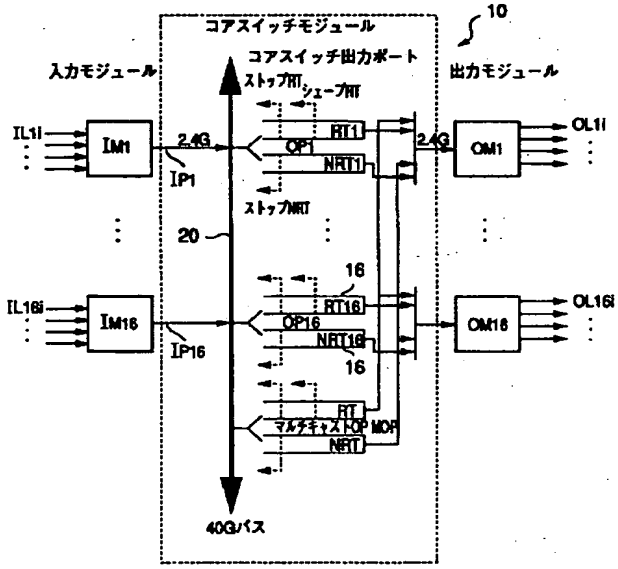
【図１７】ＣＢＲフロー及びＵＢＲフローについてＤＲＣレートの収束を示す、図１６のモデルで実施されたシミュレーションで収集されたデータのグラフである。

【図１８】コアスイッチの２つの出力ポートに与えられる３つのセルストリームフローを示す図である。

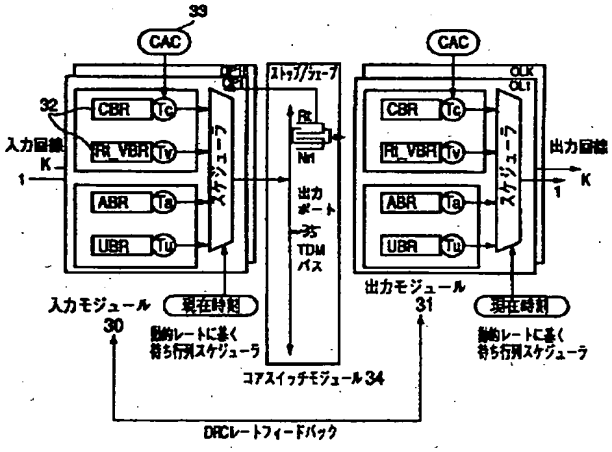
【符号の説明】

１０、３４　コアスイッチモジュール
２０、３５　ＴＤＭバス
３０　入力モジュール
３２　バッファ
３３　ＣＡＣ
１００　レジスタ
１１０　コンパレータ
２００　カウンタ
１３０　仮想キューＶＱｉｊ
１３５，１４５　ブロック
１５０　計算エンジン

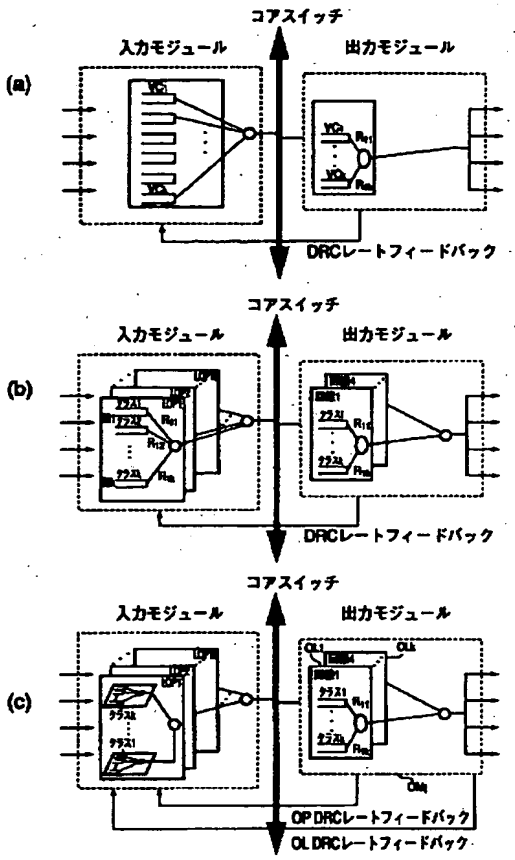
【図１】



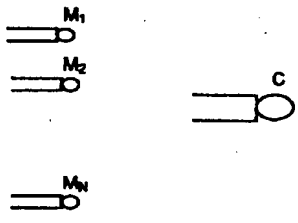
【図３】



【図２】

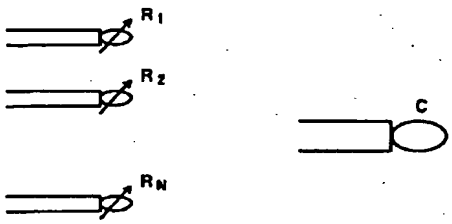


【図４】



$M_1+M_2+\cdots+M_N<C$

【図５】

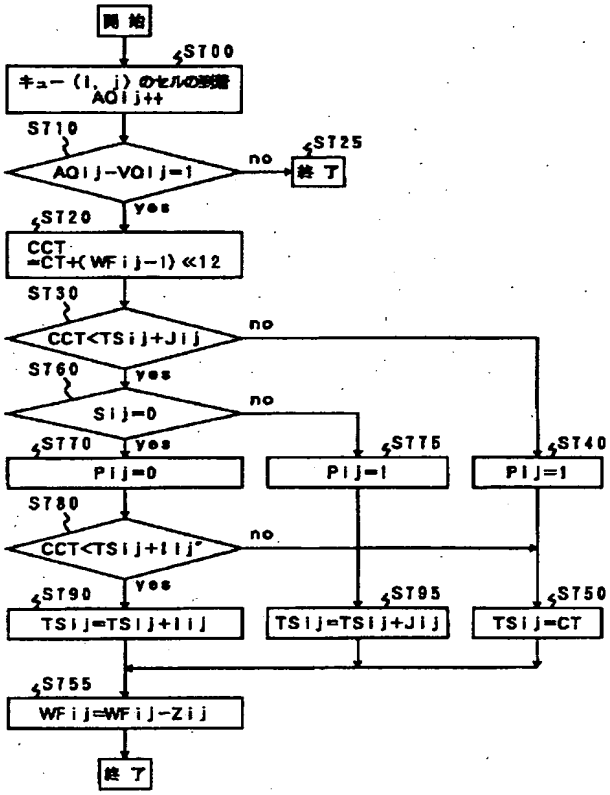


$R_1=M_1+E_1$

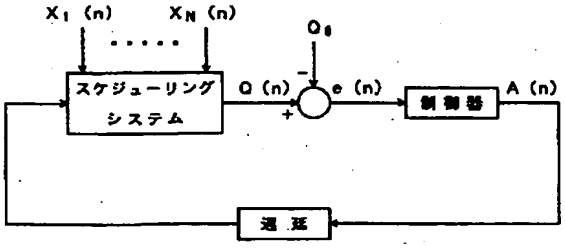
$M_1+M_2+\cdots+M_N<C$

DRCスケジューリング

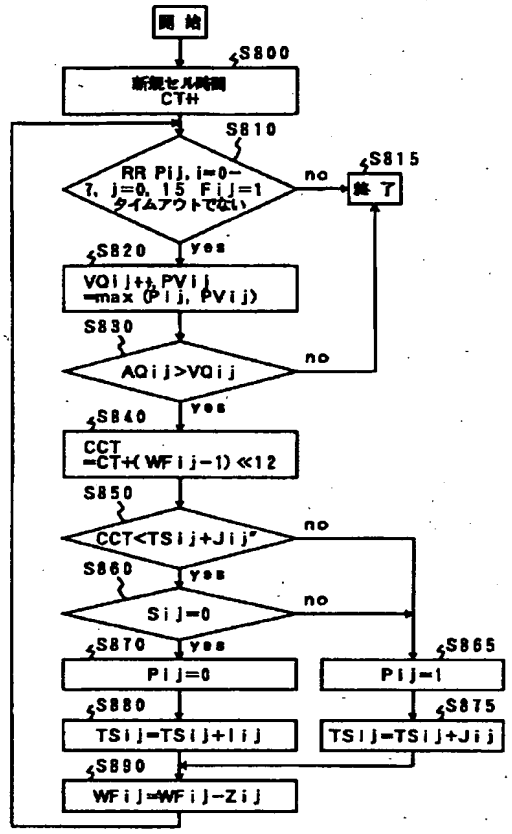
【図７】



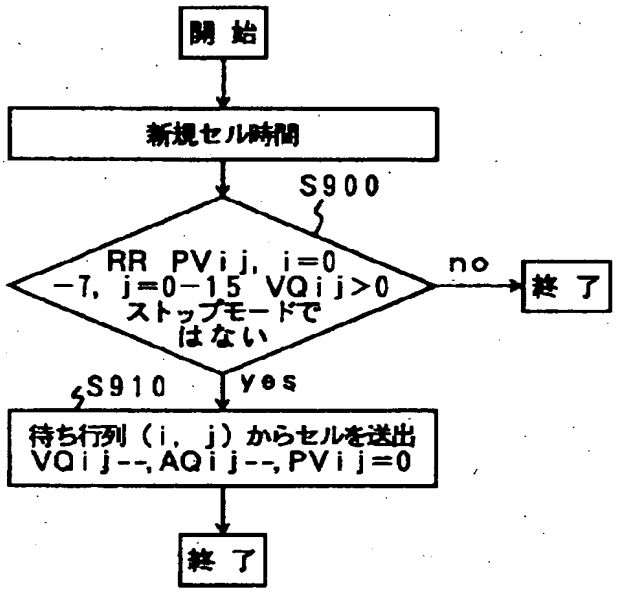
【図６】



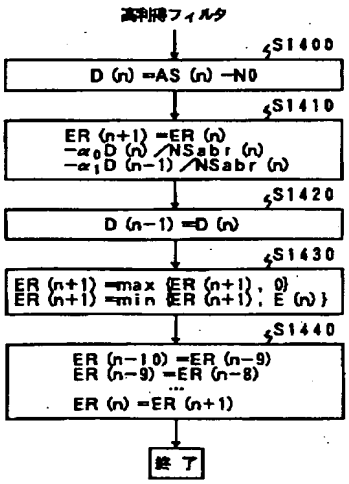
【図８】


開始
S800
新規セル時間 CT+
S810
RR Pij, i=0−7, j=0, 15 Fij=1 タイムアウトでない
no
S815
終了
yes
S820
VQij+, PVij =max (Pij, PVij)
S830
AQij>VQij
no
yes
S840
CCT =CT+(WFij−1)≪12
S850
CCT<TSij+Jij
no
yes
S860
Sij=0
no
yes
S870
Pij=0
S865
Pij=1
S880
TSij=TSij+Iij
S875
TSij=TSij+Jij
S890
WFij=WFij−Zij

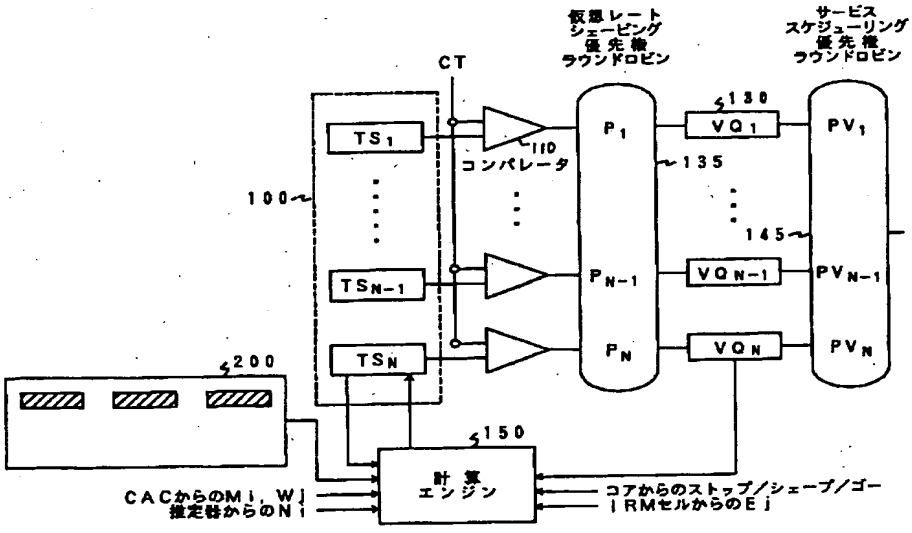
【図９】


開始
新規セル時間
S900
RR PVij, i=0−7, j=0−15 VQij>0 ストップモードでない
no
終了
yes
S910
待ち行列（i, j）からセルを送出 VQij−−, AQij−−, PVij=0
終了

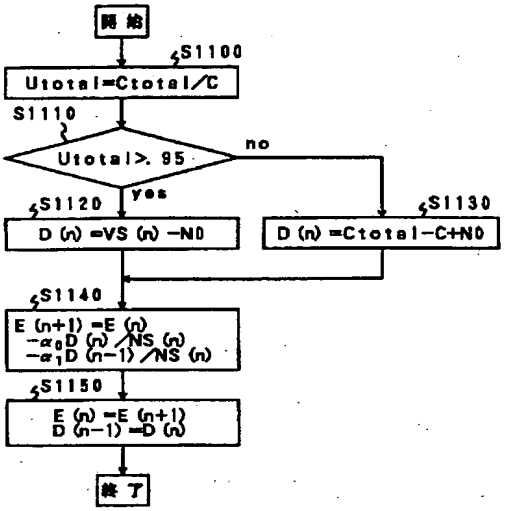
【図１４】


高帯域フィルタ
S1400
D (n) =AS (n) −N0
S1410
ER (n+1) =ER (n) −α0D (n) /NSabr (n) −α1D (n−1) /NSabr (n)
S1420
D (n−1) =D (n)
S1430
ER (n+1) =max (ER (n+1), 0) ER (n+1) =min (ER (n+1), E (n))
S1440
ER (n−10) =ER (n−9) ER (n−9) =ER (n−8) … ER (n) =ER (n+1)
終了

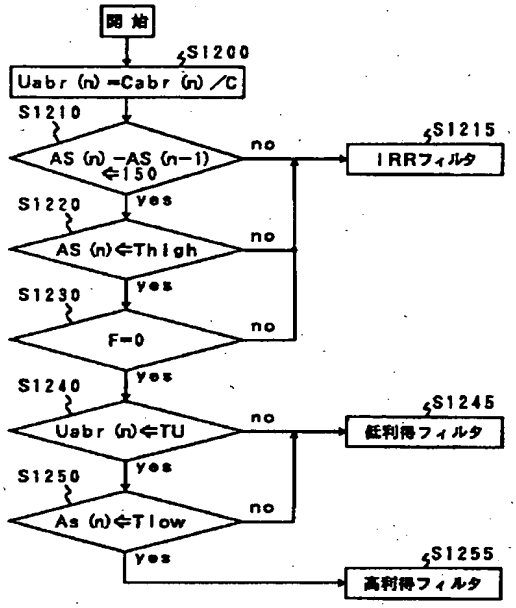
【図１０】


仮想レートシェーピング優先順ラウンドロビン
サービススケジューリング優先順ラウンドロビン
CT
TS1
TSN−1
TSN
100
110
コンパレータ
P1
PN−1
PN
130
VQ1
135
VQN−1
145
VQN
PV1
PVN−1
PVN
200
150
計算エンジン
CACからのMi, Wi 指定器からのNi
コアからのストップ/シェープ/ゴー iRMセルからのEi

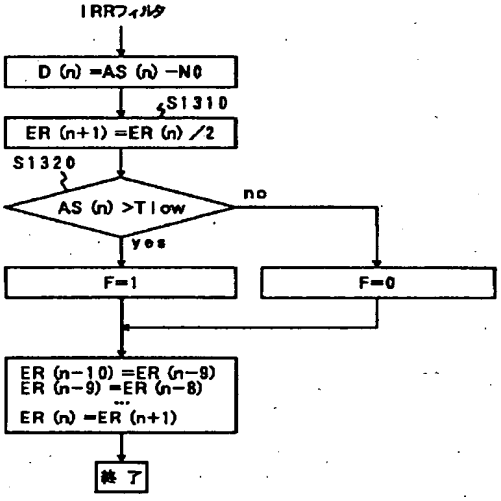
【図１１】



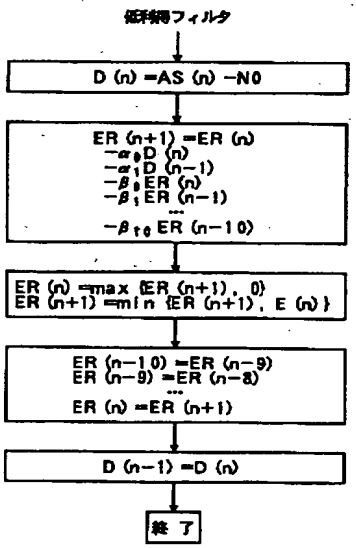
【図１２】



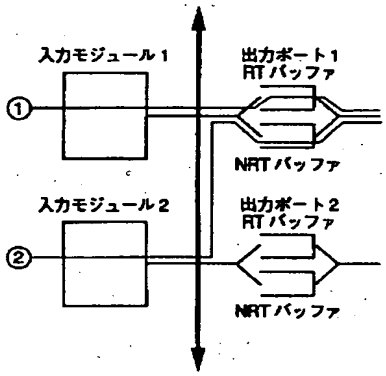
【図１３】



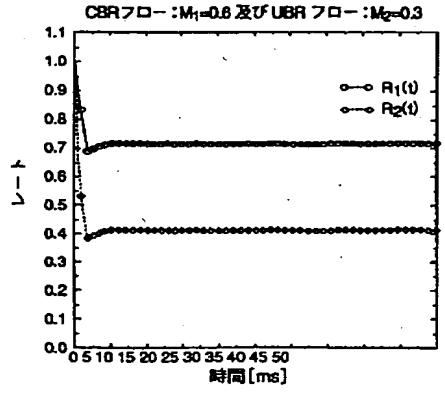
【図１５】



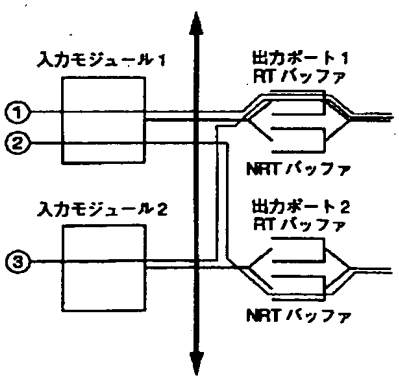
【図１６】



【図１７】



【図１８】



フロントページの続き

(72)発明者　ブライアン　マーク
アメリカ合衆国，ニュージャージー　08540，プリンストン，４　インディペンデンス　ウエイ，エヌ・イー・シー・ユー・エス・エー・インク内

(72)発明者　ゴパラクリシナン　ラママーシー
アメリカ合衆国，ニュージャージー　08540，プリンストン，４　インディペンデンス　ウエイ，エヌ・イー・シー・ユー・エス・エー・インク内